Hindawi
Journal of Theoretical Social Psychology
Volume 2022, Article ID 5118325, 15 pages
https://doi.org/10.1155/2022/5118325

Research Article

A Theory of Reliance on Individuating Information and Stereotypes in Implicit Judgments of Individuals and Social Groups

Rachel S. Rubinstein,[1] **Lee Jussim**,[2] **Bryan Loh**,[3] **and Megan Buraus**[3]

[1]*Department of Psychology, Towson University, 8000 York Road, Towson, Maryland 21252, USA*
[2]*Department of Psychology, Rutgers University—New Brunswick, 53 Avenue E, Piscataway, NJ 08854, USA*
[3]*Graduate School of Applied Psychology, Rutgers University—New Brunswick, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA*

Correspondence should be addressed to Rachel S. Rubinstein; rrubinstein@towson.edu

Received 12 March 2022; Revised 29 April 2022; Accepted 13 May 2022; Published 27 August 2022

Academic Editor: Richard Crisp

We propose a theory of (a) reliance on stereotypes and individuating information in implicit person perception and (b) the relationship between individuation in implicit person perception and shifts in implicit group stereotypes. The present research preliminarily tested this theory by assessing whether individuating information or stereotypes take primacy in implicit judgments of individuals under circumstances specified by our model and then testing the malleability of implicit group stereotypes in the presence of the same (or additional) counterstereotypic individuating information. Studies 1 and 2 conceptually replicated previous research by examining the effects of stereotype-inconsistent and stereotype-consistent individuating information on implicit stereotype-relevant judgments of individuals. Both studies showed that stereotypic implicit judgments of individuals made in the absence of individuating information were reversed when the individuals were portrayed as stereotype-inconsistent and were strengthened when targets were portrayed as stereotype-consistent (though in Study 2 this strengthening was descriptive rather than inferential). Studies 3 and 4 examined whether the strong effects of individuating information found in studies 1 and 2 extended to the social groups to which the individuals belonged. Even in the presence of up to eight counterstereotypic exemplars, there was no evidence of significant shifts in group stereotypes. Thus, the data showed that the shifts in implicit judgments that were caused by individuating information did not generalize to stereotypes of the social groups to which the individuals belong. Finally, we propose modifications to our theory that include potential reasons for this lack of generalization that we invite future research to explore.

1. Introduction

When perceivers implicitly (i.e., indirectly) judge others, to what extent are these judgments based on preexisting stereotypes (i.e., beliefs about social groups and their individual members [1])? Do these implicit beliefs dominate implicit social judgments (i.e., beliefs about or evaluations of social targets that are measured indirectly), or do perceivers take into account relevant, valid information about people?

We propose a theory (see Figure 1) that addresses (a) reliance on individuating information and stereotypes in implicit person perception and (b) the relationship between individuation in implicit person perception and shifts in implicit group stereotypes. According to this theory, individuating information takes primacy over stereotypes in implicit person perceptions under certain circumstances, and changes in implicit judgments of counterstereotypic exemplars may potentially lead to changes in implicit group stereotypes. One potential moderator of this possible generalization effect tested in the present research was the number of counterexemplars to which perceivers were exposed. In the general discussion, we propose a revision to

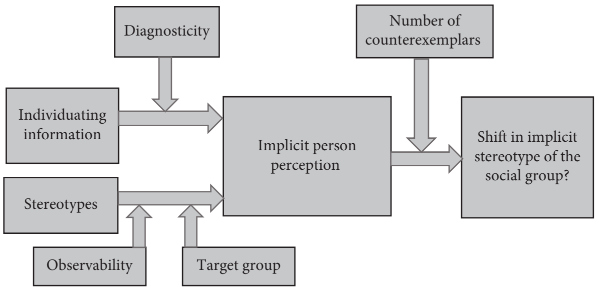
Diagnosticity
Number of counterexemplars
Individuating information
Implicit person perception
Shift in implicit stereotype of the social group?
Stereotypes
Observability
Target group

Figure 1: Theoretical model of (a) moderators of stereotype and individuating information effects in implicit person perception and (b) the relationship between individuation in implicit person perception and shifts in implicit stereotypes. Note. According to our model, individuating information only takes primacy in implicit person perception when the following three criteria are met: (a) the individuating information is highly diagnostic, (b) the stereotypes are unobservable (i.e., formed with little or no inference [2]), and (c) target groups are not gender groups. When individuating information takes primacy in implicit person perception, this may (or may not) lead to change in the stereotype of the social group (we specified competing hypotheses in this domain), and one potential moderator of this possible relationship is the number of counterexemplars that are provided (more details provided in text).

our model to incorporate potential processes underlying our results that we invite future research to explore.

In the present research, we preliminarily tested this theory with four studies. Two of these studies (studies 1 and 2) examined whether perceivers rely more heavily on individuating information or on stereotypes in implicit person perception under the circumstances addressed by our theory. Two additional studies (studies 3 and 4) tested whether group stereotypes are shifted in the presence of counterstereotypic exemplars and whether this potential effect is moderated by the number of exemplars provided (Study 4).

1.1. Do Implicit Evaluations Incorporate Valid Information from the Social Environment?

1.1.1. Theoretical Perspectives. In several dual-process theories, a distinction is drawn between slow-learning, associative "system 1" processing and rule-based, fast-learning "system 2" processing [3–5] (see also [6–8]). Some have applied this idea to the domain of social cognition, proposing that, for the most part, implicit evaluations are system 1 processes [3, 5, 9], whereas explicit evaluations are system 2 processes [3, 5]. Because system 1 processing is slow learning in nature and system 2 processing is fast learning, these theories are consistent with the notion that implicit evaluations should not readily incorporate relevant information in the social environment, whereas explicit evaluations should be sensitive to relevant information in the social environment.

In contrast, other theories are more consistent with the possibility that implicit evaluations can be rapidly updated [10], or at least allow for their revision under particular circumstances [6, 7]. According to the associative-propositional evaluation model (APE model [6, 7]), instead of rigidly distinct associative and propositional processes, these processes are indeed separate but can interact to result in similar implicit and explicit evaluations that both are sensitive to new information. According to propositional models of evaluations (e.g., [10]), there are no separate associative and propositional systems at all; instead, a single propositional process underlies all evaluations, though different conditions of automaticity may bring about dissociations between implicit and explicit evaluations [10] (see also [11]).

One dual-process theory—the APE model [6, 7]—puts forth circumstances under which counterexemplars are expected to influence group implicit evaluations. In particular, when counterinformation is affirmed (as opposed to negated), this should shift implicit group evaluations. However, while the APE model focuses on the associative vs. propositional nature of these changes, it does not connect individual-level perceptions with group-level perceptions. The present theory is the first of which we are aware that integrates literature investigating reliance on stereotypes vs. individuating information in implicit person perception with that exploring the effects of counterexemplars on implicit group stereotypes. In the revision to the model discussed later, we also offer an empirically testable, tentative social psychological account for this connection (or lack thereof).

Moreover, none of the extant theoretical accounts of implicit evaluations specify characteristics of stereotypes or individuating information that determine when perceivers rely on each source of information in implicit person perception. The present theory also addresses this theoretical gap.

1.1.2. Empirical Evidence. Within the domain of previous empirical literature that has investigated the sensitivity of implicit judgments to social information, two subsets of this research pertain most closely to the present theory. These are (a) studies that have examined reliance on individuating information and social category information in implicit judgments of individual targets and (b) studies that have investigated the revision of implicit group stereotypes in the presence of counterstereotypic exemplars.

(1) Reliance on Individuating Information and Social Category Information in Implicit Judgments of Individuals. Research examining reliance on social category information and individuating information in implicit stereotypic or attitudinal judgments of individuals has yielded mixed results. Some have shown that social category information dominates such judgments ([2, 12], studies 1 and 3). Other evidence has suggested that social category information partially influences such judgments ([2], Study 4 [13]), yet other studies have found that the influence of social category information on implicit judgments of individuals is eliminated by diagnostic (i.e., relevant) individuating information ([2], Study 2 [14], Study 3 [15], studies 1 and 2).

Despite the seemingly mixed nature of this evidence, there are several moderators of individuating information effects that have emerged from previous research. This allows us to specify in our theory conditions under which individuating information is predicted to take primacy over stereotypes in implicit person perception (or, conversely, conditions under which stereotypes are predicted to influence implicit judgments). Research investigating the effects of social category information and individuating information on attitudes in implicit person perception [12] was not included in developing these theoretical predictions because attitudes and stereotypes are distinct processes.

First, individuating information must be highly diagnostic to take primacy in implicit person perception. Somewhat diagnostic information does not have this effect [15].

Second, when stereotypes are observable (i.e., formed with minimal or no inference necessary), these stereotypes take primacy over highly diagnostic individuating information in implicit person perception [2]. For instance, the stereotype that *elderly people have poor posture* [16] is directly observed in the environment with little to no additional inference necessary, whereas the stereotype that *elderly people are lonely* [16] involves inferring a trait from behaviors or other direct observations. Thus, our theoretical perspective predicts that individuating information will take primacy in implicit judgments relevant to the stereotype about loneliness and that stereotypes will take primacy in implicit judgments relevant to stereotype about posture.

Finally, both observable and unobservable gender stereotypes tend to continue to influence implicit person perception despite highly diagnostic individuating information [2, 13]. In contrast, unobservable racial stereotypes in implicit person perception were eliminated by individuating information that portrayed two targets of different racial backgrounds as equal on the stereotyped trait dimension [14, 15] and were reversed by counterstereotypic trait information ([2], Study 2). Thus, prior research led us to predict that individuating information would take primacy over stereotypes when (a) individuating information is highly diagnostic, (b) stereotypes are unobservable, and (c) the target groups are not gender groups. When these conditions are not met, we propose that stereotypes will exert at least an equal effect as will individuating information in implicit judgments of individuals. Figure 1 delineates the proposed moderating effect of diagnosticity of individuating information on reliance on this information in implicit person perception. It also shows the hypothesized moderating effects of (a) observability of the stereotype and (b) whether the stereotype relates to gender on the effects of the stereotype on implicit person perception.

(2) Malleability of Implicit Group Evaluations. Although previous research investigating the effects of counterexemplars on implicit group *prejudice* has consistently demonstrated the effectiveness of such interventions in shifting implicit preferences (e.g., [17–22]), the results of research that has assessed the effectiveness of counterstereotypic exemplars in reducing implicit group *stereotypes* are more heterogeneous. The obtained effects seem to depend in part upon the target group. Relevant research that has employed racial target groups has found that implicit racial group stereotypes do shift in response to counterstereotypic exemplars [22–24]. However, studies assessing the malleability of group gender stereotypes given counterstereotypic exemplars find mixed results; sometimes these stereotypes are revised [25, 26], but other times they are not, particularly in the domain of gender-STEM stereotypes (e.g., [27–29]).

Research that has employed racial target groups and assessed the effects on implicit stereotypes of these groups of exposure to counterstereotypic exemplars is limited by the fact that, to our knowledge, these studies have universally employed famous counterstereotypic exemplars [22–24], which, to our awareness, leaves the effects of novel exemplars on implicit racial group stereotypes unexplored. Thus, to our awareness, there is currently a lack of experimental research examining the effects of novel counterstereotypic exemplars on implicit racial group stereotypes. Although this question has been examined in the domain of gender stereotypes [29], it is possible that different findings would have emerged if racial groups had been investigated. This is especially true given previous findings that gender stereotypes in implicit *person perception* were not fully responsive to individuating information ([2], Study 4 [13]), whereas racial stereotypes in implicit person perception were fully responsive to individuating information ([2], Study 2 [14, 15]).

Moreover, discrepant findings may emerge for novel vs. famous exemplars because associations involving famous exemplars are likely reinforced and rehearsed to a far greater extent than associations involving novel exemplars. Thus, despite consistent findings of shifts in implicit racial group stereotypes in response to counterexemplars [22–24], these effects may not emerge when novel exemplars are provided. The lack of previous research examining the effects of novel counterstereotypic exemplars on implicit group stereotypes led us to specify competing hypotheses in this domain in the present theory. The additional distinction between novel and famous counterexemplars is important because, in daily life, perceivers frequently meet new members of various social categories. Indeed, arguably, they do so more frequently than they are exposed to famous counterexemplars.

1.2. Theoretical Limitations of Previous Research.

One theoretical limitation of previous theoretical frameworks is that no previous theory of which we are aware has systematically specified conditions under which individuating information is not expected to take primacy in implicit person perception. In addition, a major theoretical limitation of the bodies of previous work that have separately addressed (a) reliance on individuating information and social category information in judgments of individuals and (b) the malleability of implicit group stereotypes in response to counterexemplars is that no previous theory of which we are aware has attempted to merge these two separate bodies of research into a unified model. The present research addressed these gaps by proposing and providing preliminary tests of a theory of (a) moderators of individuating information and stereotype effects in implicit person perception and (b) the psychological relationship between individuation in implicit judgments of individuals and changes in group stereotypes in response to counterexemplars. The present research tests one potential moderator of this latter relationship, and we later specify testable hypotheses regarding potential social psychological processes underlying our results for future research to possibly explore.

1.3. The Present Research. We performed four studies to preliminarily test a theory of (a) reliance on individuating information and stereotypes in implicit person perception and (b) the psychological relationship between individuation in implicit person perception and the effects of novel counterexemplars on implicit group stereotypes and a potential moderator of these possible effects. Studies 1 and 2 were conceptual replications of previous research demonstrating full responsiveness of unobservable stereotypes irrelevant to gender to highly diagnostic individuating information ([2], Study 2 [14, 15]). They addressed the portion of our theory specifying circumstances under which we expect individuating information to take primacy in implicit person perception. In particular, they addressed the questions as follows: (a) in the presence of highly diagnostic counterstereotypic individuating information, are implicit judgments of individual members of stereotyped groups opposite in direction relative to judgments of these same individuals in the absence of individuating information; and (b) are implicit judgments of individual members of stereotyped groups more strongly stereotypic in the presence of highly diagnostic stereotype-consistent individuating information than they are in the absence of individuating information? They tested these questions in the domain of unobservable stereotypes that were unrelated to gender to provide preliminary tests of the implicit person perception portion of our theory. Studies 3 and 4 further tested our theory by addressing the questions as follows: (c) do counterstereotypic exemplars affect implicit group stereotypes in the same manner as stereotype-inconsistent individuating information influences implicit person perception (Study 3) and (d) do increases in the number of counterstereotypic exemplars cause corresponding decreases in implicit group stereotype application (Study 4)?

We would like to emphasize that the present research did not comprehensively test our theory. Thus, the studies reported below should be considered preliminary evidence that can provide a springboard for future research.

1.4. Hypotheses

Hypothesis 1: implicit race stereotype-relevant judgments of individual members of stereotyped groups will fully incorporate highly diagnostic counterstereotypic individuating information; in the presence of counterstereotypic individuating information, such judgments will be opposite in direction as implicit judgments of these individuals made in the absence of individuating information (studies 1, 2, and 3).

Hypothesis 2: these judgments also will fully incorporate stereotype-consistent individuating information; in the presence of stereotype-consistent individuating information, implicit race stereotype-relevant judgments of individuals will be more strongly stereotypic than they will be in its absence (studies 1 and 2).

Due to the current lack of research examining the effects of novel counterexemplars on implicit group racial stereotypes and the mixed nature of evidence regarding the effects of counterexemplars on implicit group stereotypes, the remaining hypotheses were designated as sets of competing hypotheses.

Hypothesis 3a: a single pair of novel counterstereotypic exemplars (one Black and one White) will affect implicit group stereotypes to a similar extent as counterstereotypic individuating information influences implicit judgments of individuals (Study 3).

Hypothesis 3b: implicit group stereotypes will be impacted less by novel counterstereotypic exemplars than implicit judgments of individuals will be influenced by counterstereotypic individuating information (Study 3).

Hypothesis 4a: increasing the number of novel counterstereotypic exemplars will cause larger shifts in implicit group stereotypes (Study 4).

Hypothesis 4b: increasing the number of novel counterstereotypic exemplars will not affect the extent to which group stereotypes shift (Study 4).

1.5. Research Overview. The present research comprised four studies that preliminarily tested our theory. As noted above, the first two studies conceptually replicated findings from previous research investigating reliance on individuating information in implicit person perception that showed that unobservable non-gender stereotypes in implicit person perception were eliminated by highly diagnostic individuating information ([2], Study 2 [14, 15]). Study 3 manipulated whether perceivers judged individual or group targets after reviewing or not reviewing individuating information. Thus, this study directly compared the magnitude of the effects of this information on implicit judgments of individuals with its effects on implicit group stereotypes. The final study tested whether increasing the number of counterstereotypic exemplars caused a greater shift in implicit group stereotypes.

2. Studies 1 and 2

The first two studies examined the extent to which perceivers' implicit judgments of individual members of stereotyped groups in the presence of stereotype-inconsistent or stereotype-consistent individuating information were opposite in direction as or more strongly stereotypic than implicit judgments of these individuals in the absence of individuating information. Participants in these studies either were provided with no information about individuals, stereotype-consistent individuating information about the two individuals, or stereotype-inconsistent individuating information about the two individuals. In all three conditions, participants then completed a racial stereotype implicit association test (IAT [30]) to assess the extent to which they associated the Black and White targets with athleticism and intelligence (Study 1) and intelligence vs. unintelligence (Study 2), which were the stereotypes that we investigated in the two studies [31, 32]. These stereotypes are all classified as unobservable stereotypes (see [2]). Two separate conceptual replications were performed because while Study 1 avoided confounding attribute valence with the targets using two positive stereotypes (see [33]), the stereotypes in Study 2 (which involved a valence contrast) had greater ecological validity.

2.1. Method. The preregistration of Study 1 analyses and data exclusions can be found at https://osf.io/7956t/?view_only=d45f741ec1f34568b44cecd748c5c711 and that for Study 2 can be found at https://osf.io/w7mpd/?view_only=b959fd647d214733b5b3c4fb97971714. Sample size, measures, and experimental manipulation were not preregistered; the preregistrations were created after data were collected but prior to data cleaning and analysis.

2.1.1. Experimental Design. The experimental design for explicit measures in both studies was a 3 (individuating information: no information vs. stereotype-consistent information vs. stereotype-inconsistent information) X 2 (target race: Black vs. White) mixed-model design. Individuating information was the between-subjects factor. The experimental design for implicit measures was a one-way (individuating information: no information vs. stereotype-consistent information vs. stereotype-inconsistent information) between-subjects design; because IAT is difference scores, they inherently incorporate the within-subjects race of target factor.

2.1.2. Participants. Power analyses were performed for all studies in the present program of research specifying an effect size of $f = .15$ to be able to detect even small effects, $\alpha = 0.05$, 80% power, and a correlation between repeated measures of $r = 0.50$. The power analysis was based on the experimental design for implicit measures because our hypotheses are related exclusively to data from implicit measures. We specified a within-between-subjects interaction for a mixed-model ANOVA in the power analysis even though the experimental design for the IAT is a one-way between-subjects design because IAT scores are difference scores, and between-subjects main effects for difference scores are statistically identical to within-between-subjects interactions in mixed-model designs. For studies 1 and 2, the power analysis showed that 111 participants were needed. Our goal with regard to sample size was to collect more data than necessary in anticipation of discarding some data, but we did not designate a specific numerical stop point for data collection; rather, we planned to stop data collection a few days after we reached the minimum sample size to provide the extra data. Data were not analyzed at any point during data collection for any of the four studies reported.

Participant characteristics and data discards from all four studies are described in Table 1. We note that the reason for exclusion of Black and biracial participants is because implicit preferences about Black targets are different for Black perceivers than for others (e.g., [35, 36]).

2.1.3. Stimuli. Stimuli varied between the three experimental conditions in both studies.

(1) No Information Condition. In this condition, participants did not receive any stimulus information; they evaluated the targets exclusively based on racially prototypical names (Jamal DeShawn Robinson for the Black target and Luke Connor Reed for the White target; racial prototypicality was established with pilot data from our previous research, reported in Tables S1–S3 in Supplemental Materials). This condition was intended to assess the extent to which perceivers relied on stereotypes when making implicit judgments of individuals in the absence of individuating information; without specific knowledge of the individuals, perceivers tend to rely on stereotypes in judgments of individuals (e.g., [37]; see [38] for a review).

(2) Stereotype-Consistent and Stereotype-Inconsistent Information Conditions. In Study 1, in these conditions, participants read the following descriptions of Jamal and Luke.

Jamal DeShawn Robinson/Luke Connor Reed is very brainy. He has a passion for organic chemistry, and in his spare time, he plays chess, solves crossword puzzles, and reads in the library.

Jamal DeShawn Robinson/Luke Connor Reed is very athletic. He runs half-marathons a few times per year. In addition to training for these races, he plays in recreational basketball and soccer leagues and lifts weights a couple of times per week.

In the stereotype-consistent information condition, Jamal was depicted as athletic and Luke as intelligent (e.g., [31, 32]). In the stereotype-inconsistent information condition, these pairings were reversed.

In Study 2, the intelligent target description was identical to that used for the intelligent target in Study 1. However, instead of the athletic target description, participants also read the following unintelligent target description.

Jamal DeShawn Robinson/Luke Connor Reed is a brainless person. He failed his examinations and dropped out of school. He dislikes reading and enjoys watching trashy TV shows.

In the stereotype-consistent information condition, Jamal was depicted as unintelligent and Luke as intelligent (e.g., [31, 32]). In the stereotype-inconsistent information condition, these depictions were reversed.

2.1.4. Measures. The main dependent measure in all four of our studies was the IAT. The logic behind the IAT is that if a perceiver harbors a given implicit association, reaction times to stimuli under conditions consistent with this association should be faster than reaction times to stimuli under conditions inconsistent with this association. In the IAT, one or two categories appear on each top corner of the computer screen and participants sort stimuli into the appropriate

Table 1: Summary of sample characteristics.

Study characteristics	Study 1	Study 2	Study 3	Study 4
Sample	Students	Students	Students	Students
Initial sample size	131	166	207	316
#Black, biracial, or race not specified students excluded	N/A (not collected)	7	15	30
#Of fast IAT responders excluded [34]	3	3	2	6
#Excluded due to failed manipulation checks	0	0	N/A	68
#Suspicious or not following instructions excluded	0	0	0	0
#Of random deletions to adhere to IRB protocol[a]	N/A	N/A	26	46
Final sample size	128	156	165	165
Age	$M = 19.03$	$M = 18.93$	$M = 18.59$	M = 20.53
Race/ethnicity	Asian/Asian American: 76	Asian/Asian American: 75	Asian/Asian American: 86	White: 91
	White: 36	White: 58	White: 55	Asian/Asian American: 45
	Latinx/Hispanic: 12	Latinx/Hispanic: 19	Latinx/Hispanic: 16	Latinx/Hispanic: 26
	Identify with another race: 4	Identify with another race: 4	Identify with another race: 8	Identify with another race: 3
Gender	Women: 82	Men: 81	Women: 120	Women: 136
	Men: 45	Women: 74	Men: 44	Men: 29
			Identify with another gender: 1	

M = mean; SD = standard deviation. [a] The researchers collected more data than the maximum specified by the IRB due to an oversight, so the IRB required data to be deleted and destroyed until the approved sample size was reached. To randomly delete data, a random number generator generated subject numbers from the data file, and the corresponding participants were deleted from the data file. For Study 3, the patterns of results were the same when the sample size was $N = 191$ as when it was $N = 165$ with one exception, which will be noted in a footnote. The randomly deleted participants came from the following conditions: 6 from the individual targets with irrelevant, nonsocial information condition; 7 from the individual targets with counterstereotypic individuating information condition; 7 from the group target with irrelevant, nonsocial information condition; and 6 from the group target with counterstereotypic information condition. For Study 4, as a result of the randomized deletions, data from a total of 17 participants from the irrelevant, nonsocial information condition; 14 participants from the 2 pairs of exemplar condition; and 15 participants from the 4 pairs of exemplar condition were removed.

category. Single categorization trials (i.e., categorization when one category appears in each corner) are practice trials. Double categorization trials (i.e., categorization when two categories appear in each corner) are critical trials. In the double categorization trials, categories are paired together in both stereotype- or attitude-consistent and inconsistent pairings. The IAT measures differences in response latencies between the two types of pairings and thus differences in strength of association between these two pairs of categories.

We administered our IAT using the IATGen app [34] for Qualtrics (all measures in the present research were administered on Qualtrics). The categories used in the Study 1 IAT were as follows: *Jamal*, *Luke*, *Intelligent*, and *Athletic*. The stimuli in the *Jamal* and *Luke* categories were the targets' first, middle, and last names. In the *Intelligent* and *Athletic* categories, words relevant to intelligence and athleticism were used as stimuli (see Supplemental Materials). Thus, the IAT measured differences in the extent to which participants implicitly associated Jamal, compared with Luke, with intelligence and athleticism. The Study 2 IAT was identical to that used in Study 1, except that *Unintelligent* was used as a category in place of *Athletic*, and the stimulus words were changed accordingly.

Explicit measures were trait ratings of targets' athleticism (Study 1) and intelligence (studies 1 and 2) on scales of 1 (e.g., *very unintelligent*) to 7 (e.g., *very intelligent*). The order in which targets were evaluated was randomized for explicit measures, but the IAT was always administered before the explicit measures. In Study 2, participants also made estimates of targets' IQs (with standard cutoff values provided as guidelines; see Supplemental Materials for all measures). Explicit measures in both studies served mainly as manipulation checks to ensure that the individuating information successfully communicated the intended trait (see Tables S4–S6 and S11–S14 in Supplemental Materials for these analyses; these analyses showed that all manipulations were successful. See Tables S9 and S10 for Study 1 cell means, SDs, and 95% CIs for explicit measures and Tables S18 and S19 for those from Study 2, and Table S15 for additional analyses on Study 1 explicit data.).

2.1.5. Procedure. In the two conditions in which participants reviewed individuating information, they were told via instructions on the computer screen that they would be presented with information about two individuals and that they would need to memorize the information to be able to complete the remainder of the study. In these conditions, after reviewing the information, participants immediately completed manipulation checks to ensure that they were attending to the information on the screen (e.g., "Which of the following is one of Jamal's hobbies?"). They next completed the IAT, the explicit measures, questions about the purpose of the experiment, and demographic items. Finally, they were debriefed and thanked for their participation.

In the no information condition, participants started the study by completing the IAT; they were not provided with information about the targets. From that point forward, the procedure was identical to the conditions in which participants were provided with individuating information.

2.2. Study 1 Results. Data, code for analyses, and select statistical computations for all four studies can be found in Supplementary Materials. Data from explicit measures in all four studies is reported in Supplemental Materials to maintain focus on the hypotheses that the present research tested, which exclusively related to implicit measures.

2.2.1. Preliminary Data Scoring and Coding. Implicit stereotype-relevant judgments on the IAT were assessed using D scores, which were computed automatically by the IATGen app for Qualtrics [39]. This program uses the scoring algorithm suggested by Greenwald et al. [34]. In Study 1 (and in Study 4), larger positive D scores indicated stronger stereotype-consistent responses—in other words, that participants more strongly associated the Black target with athleticism and the White target with intelligence. Larger negative D scores indicated stronger stereotype-inconsistent responses—that participants more strongly associated the White target with athleticism and the Black target with intelligence. See Table S8 in Supplementary Materials for Study 1 intervariable correlations.

2.2.2. Implicit Judgments. We found stereotype-consistent or stereotype-inconsistent implicit judgments in all three individuating information conditions. D scores in the no information condition were significantly greater than 0(preregistered comparisons of D to 0.15 or −0.15 due to some evidence suggesting that IAT scores are right-biased [40] and are reported in Supplemental Materials, Tables S7, S16, S21, and S28), $D_{\mathrm{NoInformation}} = 0.17$, SD = 0.36, 95% CI = (0.06, 0.28), $t\ (42) = 3.16$, $p = 0.003$ (across all four studies, effect sizes are not reported for single-sample t-tests for D scores because the computation of D so closely resembles the computation of Cohen's d). The same was true for D scores in the stereotype-consistent information condition, $D_{\mathrm{StereotypeConsistent}} = 0.46$, SD = 0.37, 95% CI = (0.35, 0.57), $t\ (43) = 8.24$, $p < 0.001$. This indicated that implicit judgments in both the no information condition and the stereotype-consistent information condition were substantially stereotypic. Hypothesis 1 predicted that counterstereotypic individuating information would reverse the direction of implicit judgments of individuals. Consistent with this hypothesis, D scores were significantly below 0 in the stereotype-inconsistent information condition, $D_{\mathrm{Stereotype\ \ Inconsistent}} = -0.24$, SD = 0.34, 95% CI = (−0.35, −0.14), $t\ (40) = -4.61$,$p < 0.001$, indicating the presence of a substantive implicit judgment in the counterstereotypic direction.

To assess the extent to which the D scores varied among the different individuating information conditions, we performed a one-way ANOVA (individuating information: no information vs. stereotype-consistent information vs. stereotype-inconsistent information). This analysis revealed that D scores differed among the individuating information conditions, $F\ (2,\ 125) = 41.39$, $p < 0.001$, $\eta^2 = 0.40$. D scores in the stereotype-inconsistent information condition were significantly lower than D scores in the no information condition, $t\ (82) = 5.46$, $p < 0.001$, $d = 1.17$ (in Studies 1 and 2, every time pairwise comparisons were used to compare data from different individuating information conditions, p values are reported after having been multiplied by 3 in accordance with Bonferroni's correction). This provided additional support for hypothesis 1. D scores also were significantly lower in the stereotype-inconsistent information condition than they were in the stereotype-consistent information condition, $t\ (83) = 9.12$, $p < 0.001$, $d = 1.97$. A pairwise comparison of mean D scores in the stereotype-consistent information condition and the no information condition indicated that D scores in the stereotype-consistent information condition were far larger than those in the no information condition, $t\ (85) = 3.63$, $p < 0.001$, $d = 0.79$. This was consistent with hypothesis 2, which predicted that implicit judgments in the presence of stereotype-consistent individuating information would be more strongly stereotypic than implicit judgments in the absence of individuating information.

2.3. Study 2 Results

2.3.1. Preliminary Data Scoring and Coding. In Study 2 (and in Study 3), larger positive D scores indicated that participants judged the Black target as more unintelligent and the White target as more intelligent. Larger negative D scores indicated that participants judged the White target as more unintelligent and the Black target as more intelligent. See Table S17 in Supplementary Materials for Study 1 intervariable correlations.

2.3.2. Implicit Judgments. We found substantial implicit stereotype-consistent or inconsistent judgments in all three individuating information conditions. In the no information condition, D scores were significantly greater than 0, $D = 0.20$, SD = 0.34, 95% CI = (0.10, 0.29), $t\ (49) = 4.13$, $p < 0.001$. The same was true in the stereotype-consistent information condition, $D_{\mathrm{StereotypeConsistent}} = 0.32$, SD = 0.37, 95% CI = (0.21, 0.42), $t\ (50) = 6.16$, $p < 0.001$. Hypothesis 1 predicted that counterstereotypic individuating information would reverse the direction of implicit judgments of individuals. In support of this hypothesis, D scores were significantly below 0 in the stereotype-inconsistent information condition, $D_{\mathrm{StereotypeInconsistent}} = -0.25$, SD = 0.39, 95% CI = (−0.35, −0.14), $t\ (54) = -4.77$, $p < 0.001$.

A one-way ANOVA (individuating information: no information vs. stereotype-consistent information vs. stereotype-inconsistent information) revealed differences in D

scores between the experimental conditions, F (2, 153) = 35.64, $p < 0.001$, $\eta^2 = 0.31$. Providing further support for hypothesis 1, D scores were significantly lower in the stereotype-inconsistent information condition than in the no information condition, t (103) = 6.27, $p < 0.001$, $d = 1.23$ (all descriptive statistics reported above). They were also lower in the stereotype-inconsistent information condition than in the stereotype-consistent information condition, t (104) = 7.71, $p < 0.001$, $d = 1.50$. Although D scores in the stereotype-consistent information condition were higher than those in the no information condition ($D_{\text{difference}} = .12$), this difference was nonsignificant, t (99) = 1.72, p = 0.27, $d = 0.34$. Therefore, implicit stereotypic judgments in the absence of individuating information were descriptively, but not inferentially, made stronger by stereotype-consistent individuating information. This provided only limited support for hypothesis 2, which predicted that implicit stereotypic judgments would become more strongly stereotypic in the presence of stereotype-consistent individuating information.

2.4. Study 1 and 2 Discussion. Hypothesis 1, which predicted that stereotypes in implicit person perception would be reversed by counterstereotypic individuating information, was supported in both studies. Hypothesis 2, which hypothesized that stereotypes in implicit person perception would be strengthened by stereotype-consistent individuating information, was fully supported in Study 1 and partially supported in Study 2. Thus, together, studies 1 and 2 provided preliminary evidence that was mostly consistent with the hypotheses relevant to the implicit person perception portion of our theory. In particular, the results of these studies support the primacy of diagnostic individuating information over unobservable stereotypes in implicit stereotype-relevant judgments of one set of non-gender target groups.

3. Study 3

Although studies 1 and 2 provided preliminary evidence supporting the predictions of our theory relevant to implicit person perception, the question of whether the observed effects generalize to the racial groups to which the individuals belong (and, thus, the remainder of the theory) was not addressed by these studies. In Study 3, we manipulated whether participants implicitly evaluated individual or group targets and whether they reviewed counterstereotypic individuating information or irrelevant, nonsocial information. These manipulations allowed us to directly assess whether the effect of individuating information on implicit judgments of individuals was equal to the effect of counterstereotypic exemplars on implicit group stereotypes. This makes Study 3 the main preliminary test of the psychological relationship between reliance on individuating information and stereotypes in implicit person perception and the effect of these same individuals on implicit group stereotypes. This study is the first of which we are aware to directly address this question.

3.1. Method. Preregistration of analyses and data exclusion procedures can be found at https://osf.io/zu3ef/?view_only=b97afb999276410c907d1abfead89ec0. Sample size, measures, and experimental manipulations were not preregistered; the study was preregistered after data were collected but prior to data cleaning and analysis.

3.1.1. Experimental Design. The experimental design for explicit measures was a 2 (target: individuals vs. racial groups) X 2 (information: irrelevant nonsocial information vs. stereotype-inconsistent individuating information) X 2 (target race: Black vs. White) mixed-model design in which target race was the within-subjects factor. The experimental design for the IAT was a 2 (target: individuals vs. racial groups) X 2 (information: irrelevant nonsocial information vs. counterstereotypic individuating information) between-subjects design. Because IAT scores are difference scores, they inherently incorporate the within-subjects target race factor from the experimental design for explicit measures.

3.1.2. Participants. A power analysis using the same parameters as did the previous studies indicated that the necessary sample size to detect a small effect in this study was $N = 128$. The stop point for data collection was determined in the same way as were those in studies 1 and 2. See Table 1 for sample characteristics and data exclusions in all four studies.

3.1.3. Stimuli. In the stereotype-inconsistent individuating information condition, participants read both of the following target descriptions (see Table S20 for Study 3 target name pilot data).

Jamal Terrell Jackson got a combined score (math + verbal) of 1580 on his SATs in high school, which was better than 99% of all other high school students. He eventually graduated from Princeton University's Physics Ph.D. program.

Eric Keith Reed received a combined SAT score (math + verbal) of 710, which was worse than 96% of all other high school students. He ended up dropping out of high school.

In the irrelevant nonsocial information condition, participants read the following statements.

Some notebooks have 80 sheets of paper, while others have 160 sheets of paper.

When photocopies are made, the paper is aligned in a particular way to make sure that the copies come out properly.

3.1.4. Measures. In the individual target condition, the IAT and explicit measures were identical to those used in Study 2 (other than changing the names to match the target descriptions). In the group target condition, the targets for the explicit measures and IAT were *Black people* and *White people* instead of the individual targets, and corresponding stimulus items in the IAT were prototypically Black and White male names that were not included in the target descriptions (see Supplemental Materials for IAT categories and stimuli and for pilot data regarding racial prototypicality of target names). The results of all explicit measures are reported in Supplemental Materials since the research hypotheses are related only to implicit measures.

3.1.5. Procedure. The procedure for this study differed slightly from that of the previous studies. In Study 3, participants started by reading information (either counterstereotypic individuating information or irrelevant nonsocial information) and then were asked to get the experimenter and repeat to the experimenter the information they had just read. This was done to ensure that they had attended to the information. If they recited the information correctly, they proceeded to the remainder of the study, but if their recall was incorrect, they were asked to try again until it was correct. After learning the information, participants completed one of two sets of dependent measures: those for the individual targets (who had the same names as the individuals in the counterstereotypic individuating information condition) or those for the group targets (*Black people* and *White people*; see Supplemental Materials for all dependent measures). They were then probed for suspicion and debriefed.

3.2. Results and Discussion. Hypothesis 1 predicted that counterstereotypic individuating information would reverse the direction of implicit stereotype-relevant judgments of individual targets relative to such judgments made in the absence of individuating information. Hypothesis 3a predicted that the counterstereotypic information would have an equal effect on implicit judgments of individuals and groups, and competing hypothesis 3b predicted that implicit group judgments would be affected to a lesser extent by the information about individuals than would implicit judgments of these same individuals.

D scores were subjected to a 2 (target: individuals vs. groups) X 2 (information: irrelevant nonsocial information vs. counterstereotypic individuating information) between-subjects ANOVA. Although the categories in the IATs differed between the two target conditions (Black people/White people in the groups condition and Jamal/Eric in the individual condition), we did not standardize the IAT scores for two reasons: (1) although the categories technically differed, they represented highly similar constructs (a Black and a White person, or Black and White people in general), and (2) when D scores are computed, the mean difference between response latencies for stereotype-inconsistent and stereotype-consistent trials for each participant is divided by the standard deviation of all of that participant's responses, and this is very similar to conventional standardization. D (See Table S22 in Supplemental Materials for Study 3 intervariable correlations).

The effect of interest was a significant target × information interaction, $F(1,\ 161) = 21.82$, $p < 0.001$, $\eta^2 = .10$ (see Table S23 for Study 3 D score main effect cell means, SDs, and 95% CIs). In support of hypothesis 1, simple-effects analysis showed that in the individual target condition, D scores were significantly lower when counterstereotypic information was provided, $M = -0.34$, SD = 0.28, 95% CI = (−0.43, −0.26), than when irrelevant nonsocial information was provided, $M = 0.19$, SD = 0.44, 95% CI = (0.05, 0.33), t (81) = 5.97, $p < 0.001$, $d = 1.45$. Providing further support for hypothesis 1, D scores in the former cell were significantly lower than 0, revealing that participants evaluated Jamal as more intelligent than Eric, t (42) = −8.06, $p < 0.001$. However, in the group target condition, there was no statistically significant difference between judgments that were made in the presence of counterstereotypic individuating information, $M = 0.20$, SD = 0.36, 95% CI = (0.09, 0.31), and irrelevant nonsocial information, $M = 0.14$, SD = 0.52, 95% CI = (−0.03, 0.31), t (80) = −0.66, p = 0.512, $d = 0.13$.

Taken together, these results preliminarily showed that although counterstereotypic individuating information led to a reversal of the direction of implicit stereotype-relevant judgments of these same individuals, there was insufficient evidence to conclude that counterstereotypic exemplars affected implicit stereotypes of the racial groups to which the exemplars belonged (see Supplemental Materials (including Table S24) for results from explicit dependent measures). This provided support to hypothesis 3b, which predicted that counterstereotypic information would affect implicit group stereotypes to a lesser extent than it would influence implicit judgments of the individuals described by the counterstereotypic information.

4. Study 4

Although Study 3 found insufficient evidence to conclude that a single pair of counterstereotypic exemplars changed implicit stereotypes in the domain of group targets, it was possible that, if multiple pairs of exemplars had been encountered, this would have been a more effective means of shifting implicit group stereotypes ([24]; cf. [28]). Study 4 tested this possibility as a moderator of the potential relationship between perceptions of counterstereotypic individuals and shifts in implicit group stereotypes.

As was the case in studies 1 and 2, we tested different racial stereotypes in Study 4 than in Study 3. In Study 4, we tested two positive stereotypes (the stereotypes that Black people are athletic and White people are intelligent); this was intended to maintain a balance in the present research of mitigating valence confounds while also testing stereotypes that have greater ecological validity.

4.1. Method. Preregistrations for most data exclusions and analyses can be found at https://osf.io/qy3nv/?view_only=be1d9303dd874de1b019e403d6ba2e15. Exceptions are noted below. Sample size, measures, and the experimental manipulation were not preregistered; this study was preregistered after data collection was complete, but prior to data cleaning and analysis.

4.1.1. Experimental Design. The experimental design for explicit measures was a 3 (information: irrelevant nonsocial information vs. two pairs of exemplars vs. four pairs of exemplars) X 2 (target race: Black vs. White) mixed-model design; target race was the within-subjects factor. The experimental design for the implicit measure for Study 4 was a one-way (information: irrelevant, nonsocial information vs. two pairs of exemplars vs. four pairs of exemplars) between-subjects design. Because IAT scores are difference scores, the experimental design for the IAT inherently incorporated the within-subjects target race factor.

4.1.2. Participants. A power analysis using the same parameters as studies 1–3 revealed that 111 participants were needed to detect a small effect on implicit measures. See Table 1 for sample characteristics and data exclusions for all four studies.

4.1.3. Stimuli. In this study, participants read either the same irrelevant, nonsocial information as did some participants in Study 3, descriptions of two Black counterstereotypic exemplars (i.e., the Black targets were depicted as highly intelligent) and two White counterstereotypic exemplars (i.e., the White targets were depicted as very athletic), or descriptions of four Black counterstereotypic exemplars and four White counterstereotypic exemplars. All exemplar descriptions are available in Supplemental Materials. A pilot test was performed to identify additional prototypically Black and White first names for Study 4 (see Table S25 in Supplemental Materials for these data).

4.1.4. Measures. The IAT was used as the measure of implicit stereotypes; as in the group target condition in Study 3, the IAT employed the target categories *Black people* and *White people*, and the stimulus items for these categories (prototypically Black and White first names) differed from all names that were presented as exemplars. The attribute categories were *Intelligent* and *Athletic*, and stimulus items for these categories were words relevant to these attributes (see Supplemental Materials for all categories and stimuli). Participants also evaluated the intelligence and athleticism of Black and White people (as groups) on scales of 1 (*very unintelligent*, *very unathletic*) to 7 (*very intelligent*, *very athletic*), and as manipulation checks, they evaluated the intelligence and athleticism of each of the individual targets (results showed that all manipulations were successful; see Tables S26 and S27 in Supplemental Materials). In addition, participants estimated the IQs of Black and White people. The results of all explicit measures are reported in Supplemental Materials since the research hypotheses are related only to implicit measures.

In this study, participants viewed up to eight target descriptions, all with first and last names. Because the IAT in this study was intended to exclusively measure implicit group stereotypes rather than stereotypes in implicit person perception, we wanted to be sure that participants did not confuse IAT stimulus names with names that they knew they had seen. Thus, we included a manipulation check to test whether participants knew that the names used as stimuli in the IAT were not included in the target descriptions. In this manipulation check, between reviewing the individuating information and taking the IAT, participants were presented with a list of first names including the first names from the target descriptions and the first names used as IAT stimuli and were asked to identify which they had seen in the target descriptions. We discarded data from the 68 participants who responded to any of these items incorrectly to ensure that we exclusively tested whether judgments of exemplars generalized to other group members. This did have the effect of making the data exclusions (and, thus, sample sizes) uneven across conditions; the more exemplars there were, the more difficult this manipulation check was. Thus, the majority of the participants discarded on the basis of this manipulation check were from the four pairs of exemplar condition. None were from the irrelevant, nonsocial information condition because these participants did not read information about specific targets.

4.1.5. Procedure. Participants completed this study online. Due to concerns that participants would not remember multiple exemplar descriptions and names without reinforcement, manipulation checks about the content of each target description (e.g., "Which of the following (activities) does Luke do to train for a marathon?") appeared after each target description. Participants were not permitted to continue without answering them correctly; they were allowed up to three attempts to do so. All participants answered all questions correctly within this allotted number of attempts. After each content manipulation check, participants were asked to rate the intelligence and athleticism of the target. After reading all of the descriptions completing all content manipulation checks, and rating each target, participants completed the name recognition manipulation check. They then completed the IAT, explicit dependent measures, and suspicion checks and were finally debriefed regarding the purpose of the experiment.

4.2. Results and Discussion. Hypothesis 4a posited that larger numbers of counterstereotypic exemplars would cause greater reductions in implicit group stereotypes. Competing hypothesis 4b posited that larger numbers of counterstereotypic exemplars would not cause greater shifts in implicit group stereotypes. We performed a one-way between-subjects ANOVA (information: irrelevant nonsocial information vs. two pairs of counterstereotypic exemplars vs. four pairs of counterstereotypic exemplars) on D scores to directly test these hypotheses (see Table S29 in Supplemental Materials for Study 4 intervariable correlations). This analysis revealed that the differences among information conditions were not statistically significant, $F(2, 162) = 0.81$, $p = 0.449$, $\eta^2 = 0.01$. A Bayes factor (BF_{10}) of 0.13 showed that the evidence in favor of the null hypothesis was moderate to strong (BF_{10} was interpreted according to the following guidelines [41, 42]: $0.33 < \mathrm{BF}_{10} < 1$ constitutes anecdotal evidence in favor of the null hypothesis, $0.10 < \mathrm{BF}_{10} < 0.33$ is moderate evidence in favor of the null hypothesis, $0.03 < \mathrm{BF}_{10} < 0.10$ is strong evidence in favor of the null hypothesis, $0.01 < \mathrm{BF}_{10} < 0.03$ is very strong evidence in favor of the null hypothesis, and $BF_{10} < 0.01$ is extreme evidence in favor of the null hypothesis). Thus, there was insufficient evidence to conclude that the means in the various information conditions differed.

5. General Discussion

We proposed a theory that had two components. First, it systematically specified circumstances under which individuating information would take primacy over stereotypes in implicit person perception and circumstances under which stereotypes would be expected to take primacy over individuating information. It also explored the nature of the relationship between individuation in implicit person perception and shifts in implicit social group stereotypes.

The studies discussed in this report provided some preliminary data to test the theory by investigating the extent to which stereotypes and individuating information influence implicit stereotype-relevant judgments of individuals and groups. We first assessed whether individuating information, indeed, takes primacy over stereotypes in person perception under the conditions specified by our model: (a) the individuating information is highly diagnostic, (b) the stereotype is unobservable, and (c) the target groups are not gender groups. We found support for the hypothesis that individuating information would take primacy under these circumstances (though the present studies did not provide a comprehensive test of this hypothesis). In studies 3 and 4, we assessed whether the effects of individuating information on implicit person perception generalized to the target groups to which the individuals belonged. We found no evidence that the effect is generalized, even in the presence of four pairs of counterexemplars (see Figure 2 for diagram of support for hypotheses).

These predictions and related preliminary data constitute the first attempt of which we are aware to systematically specify multiple conditions that must be met for individuating information to take primacy in implicit person perception; previous research [2, 14, 15] has focused on individual moderators that the present theory integrated into a unified prediction.

In addition, the present theory was the first to our knowledge to connect implicit individual and group implicit judgments in the presence of counterinformation. Studies 3 and 4 also, to our knowledge, comprise the first studies to examine the effects of novel counterstereotypic exemplars on implicit racial stereotypes; previous research that has shown the effectiveness of counterexemplars in reducing implicit racial group stereotypes has uniformly relied on well-known exemplars [22–24]. Because it is common to encounter novel members of social groups in everyday life, and these individuals may differ from group stereotypes, this was a meaningful distinction.

5.1. Accounting for the Inefficacy of Counterexemplars at Shifting Implicit Group Stereotypes

5.1.1. Theoretical Explanations. The finding that individual- and group-level implicit judgments were discrepant even in the presence of multiple counterexemplars invites testable questions regarding phenomena that may have caused these results. Given the outcome of our competing hypotheses, we propose a revision to the theoretical model presented in Figures 1 and 2. This revision is proposed as a preliminary effort to guide future research in attempting to understand why individual counterstereotypic exemplars had no significant effect on the group stereotype.

The revised model (Figure 3) proposes that counterexemplars may trigger one or more of three classic phenomena in social psychology that might serve as stereotype-protective processes that insulate implicit group stereotypes from individuating information effects (though this proposal does not preclude other stereotype-protective processes, nor does it preclude the presence of moderators of potential effects of counterexemplars on implicit group stereotypes; see [40] for a review). The first of these phenomena is motivated ingroup processes. In particular, according to social identity theory (e.g., [43–46]), the self-enhancement-driven desire to favor the ingroup when comparing the ingroup to the outgroup causes both social category effects and stereotypic judgments that favor the ingroup over the outgroup to strengthen [47]. Thus, the downstream effects of self-enhancement motives may cause perceivers to continue to rely on stereotypes despite the presence of counterexemplars (see also [48]), especially when counterexemplars involve an ingroup target who is characterized negatively and an outgroup target who is characterized positively (as was the case in two of the present studies).

We also believe that confirmation bias (e.g., [49]; see [50] for a review) may play a role in the continued influence of stereotypes in implicit group judgments. Perceivers tend to notice and remember behaviors that they expect on the basis of stereotypes (for reviews, see [50, 51]). Thus, if perceivers have stereotypes that they believe are true (regardless of their actual truth value), they may interpret individuating information in ways that are consistent with their preexisting stereotypes, which would only serve to further reinforce these stereotypes. For instance, Darley and Gross [52] found that perceivers interpreted behavioral target information in ways that were consistent with their previous stereotypes; thus, new behavioral information served to reinforce previous beliefs. Moreover, according to the principles of confirmation bias, perceivers tend to ignore information that is inconsistent with the preexisting belief. Thus, if individuating information is counterstereotypic, perceivers may ignore it when making group judgments. This is consistent with the ineffectiveness of counterexemplars at shifting implicit group stereotypes that were demonstrated in the present research.

Finally, subtyping—a phenomenon in which stereotype-disconfirming exemplars are viewed as "exceptions to the rule" rather than being integrated into the group stereotype (e.g., [53])—may also serve to insulate implicit group stereotypes from the effects of counterexemplars. This is especially true because, in our studies, the counterexemplars were grouped together rather than dispersed among other exemplars and because they all were highly atypical of the stereotype; these are circumstances that are germane to subtyping (e.g., [54, 55]; see [56] for a review).

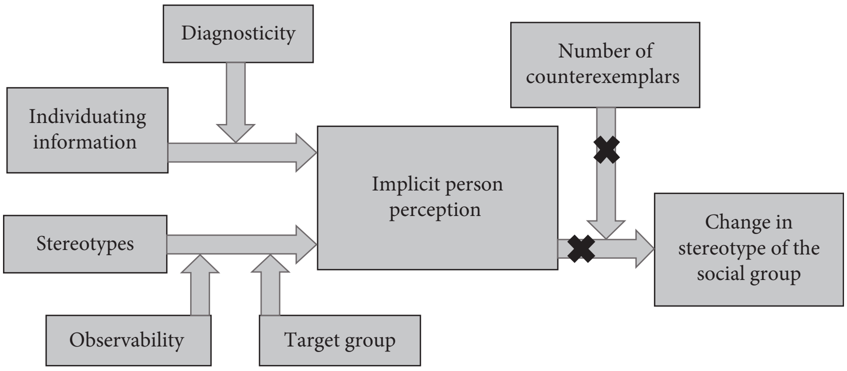
Diagnosticity
Number of counterexemplars
Individuating information
Implicit person perception
Stereotypes
Change in stereotype of the social group
Observability
Target group

Figure 2: Diagram of support for hypotheses. Note. Black *X*s indicate null effects. The present research was only capable of providing preliminary support or lack of support for the hypotheses.

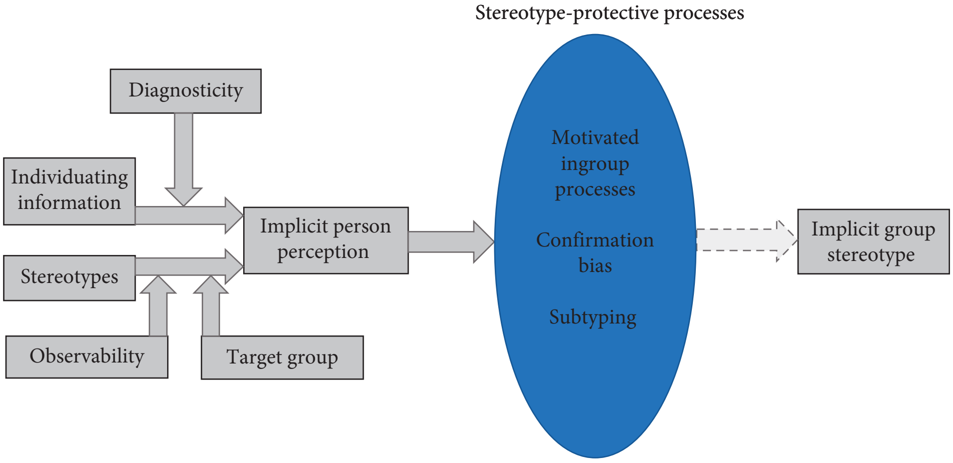
Stereotype-protective processes
Diagnosticity
Individuating information
Implicit person perception
Stereotypes
Observability
Target group
Motivated ingroup processes
Confirmation bias
Subtyping
Implicit group stereotype

Figure 3: Revised model including proposed stereotype-protective processes. Lighter arrow with a dashed outline represents the hypothesis that the stereotype-protective processes weaken the influence of implicit judgments of individuals on the implicit group stereotype. These processes were not tested in the present research; instead, they represent avenues for future research.

5.1.2. Testable Predictions. If these phenomena, indeed, can account for a lack of generalization of the exemplars to the stereotype of the social group, this would be supported by future research that seeks to diminish these influences and then test whether implicit group stereotypes are reduced by the counterexemplars. With regard to motivated group processes, if self-enhancement motives drive stereotype-protective effects, then self-affirmation should promote the effects of counterexemplars on implicit group stereotypes; indeed, self-affirmation has been established as an effective means of reducing prejudice (see [57] for a review).

Likewise, evidence has shown that confirmation bias can be reduced by strategies that encourage perceivers to consider alternative perspectives (e.g., [58, 59]). Thus, in the case of stereotypes that are inaccurate, exposure to factual information that disconfirms the stereotype (rather than counterexemplars) may enhance the effects of counterexemplars on implicit group stereotypes. This is especially true because there is no evidence that perceivers are unaware of their implicit beliefs and attitudes [60].

Finally, as noted above, subtyping is most likely to occur when counterexemplars are concentrated and blocked rather than dispersed among other exemplars and when the counterexemplars are extreme (see [56] for a review). Thus, if subtyping insulates implicit group stereotypes from the effects of counterexemplars, then decreasing the likelihood of subtyping by dispersing moderate exemplars should increase the likelihood that stereotypes will be responsive to counterexemplars.

One additional question is whether one of these phenomena or whether more than one of these phenomena must be suppressed to allow individuating information effects to shift the implicit group stereotype. We are agnostic with regard to this issue; it is an open question.

5.2. Reconciling Discrepancies. Past research that has investigated the effects of counterstereotypic exemplars on implicit racial group stereotypes [22–24] showed that implicit racial group stereotypes were shifted by counterstereotypic exemplars. These findings are inconsistent with those of studies 3 and 4 in the present research. This discrepancy may be attributable to the fact that the present research employed novel exemplars, whereas previous research relied on well-known exemplars. It is possible that, because the links between the traits and the individuals are more well-rehearsed for known exemplars due to previous exposure, they are more effective in reducing implicit stereotypes. Systematically addressing the role of familiar versus unfamiliar counterstereotypic exemplars in changing implicit stereotypes could be a fruitful area for future research.

Moreover, the results of the present research are inconsistent with previous results showing that implicit gender stereotypes are reduced by counterexemplars. Some research on the effects of counterexemplars on implicit gender stereotypes used famous exemplars [26], as did the aforementioned research finding effects of counterexemplars on implicit race stereotypes; thus, the present research may have failed to replicate these effects because perceivers considered novel exemplars in the present research.

Other research investigating the effects of counterexemplars on implicit gender stereotypes had participants from detailed mental images of "strong women" [25]; they were told to consider such matters as what her hobbies and activities are, her capabilities, what she is like, and why she is considered strong. In the present research, participants considered fewer dimensions of the trait; they merely read about some behaviors relevant to the trait. Thus, the manipulation from that previous program of research [25] arguably was stronger than that employed in the present research. Moreover, in the previous research, given that participants were told to "imagine" an exemplar, the exemplar varied among perceivers, whereas in the present research, the information about the exemplars was provided and held constant within each experimental condition. It is not possible to know what types of exemplars participants in the previous research imagined; it is possible that they imagined women who were in some way less likely to engage stereotype-protective processes than were the targets described in the present research (e.g., perhaps they imagined more moderate exemplars, who are less likely to be subtyped than extreme exemplars [56]).

5.3. Limitations and Future Directions. One limitation of the present research is that the data did not provide a comprehensive test of our theory. First, the empirical tests were restricted to a limited range of stereotypes, individuating information, target groups, and manipulations. For instance, in studies 1 and 2, our research did not test the effects of highly diagnostic individuating information on gender stereotypes nor on observable stereotypes to empirically demonstrate reliance on stereotypes under such conditions; only past literature ([2], studies 1, 3, and 4; [13]) has tested this. Similarly, we did not test the effects of individuating information that was not highly diagnostic to empirically test its weaker influence on implicit person perception; only past literature has ([15], Study 1).

Similarly, although we propose further developments to our theory in the general discussion that may help to explain our results, these were not specified in our original model due to the competing nature of our original hypotheses regarding the effects of counterexemplars on implicit group stereotypes. Thus, the process portion of our model remains untested. We invite future research to test the proposed processes.

In addition, only one moderator of a potential relationship between individuation effects in implicit person perception and the effects of counterexemplars on implicit group stereotypes was explored: the number of counterexemplars. There are many potential others (see [40] for a review). Thus, we hope that the present theoretical framework will serve as a springboard for future investigations of this question.

Further, the only measure of implicit stereotypes that we used was the IAT, and the interpretation of IAT scores has been the subject of debate (e.g., [61]). Future research should seek to replicate the present findings using a different implicit measure.

Finally, the present research was conducted at a politically liberal and racially diverse public university. It is possible that results would have differed in more conservative populations or in those that have less exposure to racially minoritized groups, so future research should address these questions using other populations.

6. Conclusion

The present research developed and tested a new theory of individuation effects in implicit person perception and of the relationship between individuation in implicit person perception and changes in implicit group stereotypes. The results across four studies preliminarily showed that strong individuating information effects on perceptions of individuals did not generalize to the social groups to which individuals belonged; perceivers relied exclusively on individuating information in implicit stereotype-relevant judgments of individuals, but exclusively on racial stereotypes in implicit judgments of the social groups. The findings of the present research were consistent with a recent, growing body of research showing that implicit judgments of individuals are consistent with valid information in the social environment (e.g., [14, 15, 62]; cf. [2], studies 1, 3, and 4 [13]). Although this is cause for optimism with regard to identifying additional avenues to mitigate social biases, the finding that this shift did not extend to the social groups to which the individuals belonged suggests that there is still much progress to be made with regard to identifying the circumstances under which implicit group stereotypes are revised. We hope that the proposed future directions discussed in our theory will provide some direction in this regard.

Data Availability

Data for all studies in this program of research are available at https://osf.io/ruw3q/?view_only=a033bd35532d4d67bd03058447e25f8b.

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Materials

Supplemental materials contain methodological information necessary for replication of the present studies that is not provided in text and analyses not reported in text. (*Supplementary Materials*)

References

[1] R. D. Ashmore and F. K. Del Boca, "Conceptual approaches to stereotypes and stereotyping," in *Cognitive Processes in Stereotyping and Intergroup Behavior*, D. L. Hamilton, Ed., pp. 1–35, Erlbaum, Hillsdale, NJ, USA, 1981.

[2] R. S. Rubinstein, L. Jussim, J. E. Bock, and B. T. Loh, "Unobservable stereotypes are more malleable than observable stereotypes in implicit person perception," *Journal of Theoretical Social Psychology*, vol. 5, no. 4, pp. 318–337, 2021.

[3] R. J. Rydell and A. R. McConnell, "Understanding implicit and explicit attitude change: a systems of reasoning analysis," *Journal of Personality and Social Psychology*, vol. 91, no. 6, pp. 995–1008, 2006.

[4] S. A. Sloman, "The empirical case for two systems of reasoning," *Psychological Bulletin*, vol. 119, no. 1, pp. 3–22, 1996.

[5] E. R. Smith and J. DeCoster, "Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems," *Personality and Social Psychology Review*, vol. 4, no. 2, pp. 108–131, 2000.

[6] B. Gawronski and G. V. Bodenhausen, "Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change," *Psychological Bulletin*, vol. 132, no. 5, pp. 692–731, 2006.

[7] B. Gawronski and G. V. Bodenhausen, "The associative-propositional evaluation model. Theory, evidence, and open questions," in *Advances in Experimental Social Psychology*, M. P. Zanna, Ed., vol. 38, pp. 59–127, 2011.

[8] T. D. Wilson, S. Lindsey, and T. Y. Schooler, "A model of dual attitudes," *Psychological Review*, vol. 107, no. 1, pp. 101–126, 2000.

[9] A. R. McConnell and R. J. Rydell, "The systems of evaluation model," in *Dual-Process Theories of the Social Mind*, J. W. Sherman, B. Gawronski, and Y. Trope, Eds., pp. 204–217, Guilford Press, New York, NY, USA, 2014.

[10] J. D. Houwer, "A propositional model of implicit evaluation," *Social and Personality Psychology Compass*, vol. 8, pp. 342–353, 2014.

[11] J. Cone, T. C. Mann, and M. J. Ferguson, "Chapter three-changing our implicit minds: how, when, and why implicit evaluations can be rapidly revised," *Advances in Experimental Social Psychology*, vol. 56, pp. 131–199, 2017.

[12] A. R. McConnell, R. J. Rydell, L. M. Strain, and D. M. Mackie, "Forming implicit and explicit attitudes toward individuals: social group association cues," *Journal of Personality and Social Psychology*, vol. 94, no. 5, pp. 792–807, 2008.

[13] J. Cao and M. R. Banaji, "The base rate principle and the fairness principle in social judgment," *Proceedings of the National Academy of Sciences*, vol. 113, no. 27, pp. 7475–7480, 2016.

[14] R. S. Rubinstein and L. Jussim, "Stimulus pairing and statement target information have equal effects on stereotype-relevant evaluations of individuals," *Journal of Theoretical Social Psychology*, vol. 3, no. 4, pp. 231–249, 2019.

[15] R. S. Rubinstein, L. Jussim, and S. T. Stevens, "Reliance on individuating information and stereotypes in implicit and explicit person perception," *Journal of Experimental Social Psychology*, vol. 75, pp. 54–70, 2018.

[16] D. F. Schmidt and S. M. Boland, "Structure of perceptions of older adults: evidence for multiple stereotypes," *Psychology and Aging*, vol. 1, no. 3, pp. 255–260, 1986.

[17] C. Columb and E. A. Plant, "Revisiting the Obama effect: exposure to Obama reduces implicit prejudice," *Journal of Experimental Social Psychology*, vol. 47, no. 2, pp. 499–501, 2011.

[18] N. Dasgupta and A. G. Greenwald, "On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals," *Journal of Personality and Social Psychology*, vol. 81, no. 5, pp. 800–814, 2001.

[19] J. A. Joy-Gaba and B. A. Nosek, "The surprisingly limited malleability of implicit racial evaluations," *Social Psychology*, vol. 41, no. 3, pp. 137–146, 2010.

[20] C. K. Lai, M. Marini, S. A. Lehr et al., "Reducing implicit racial preferences: I. A comparative investigation of 17 interventions," *Journal of Experimental Psychology: General*, vol. 143, no. 4, pp. 1765–1785, 2014.

[21] C. K. Lai, A. L. Skinner, E. Cooley et al., "Reducing implicit racial preferences: II. Intervention effectiveness across time," *Journal of Experimental Psychology: General*, vol. 145, no. 8, pp. 1001–1016, 2016.

[22] E. A. Plant, P. G. Devine, W. T. Cox et al., "The Obama effect: decreasing implicit prejudice and stereotyping," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 961–964, 2009.

[23] C. Columb and E. A. Plant, "The Obama effect six years later: the effect of exposure to Obama on implicit anti-black evaluative bias and implicit racial stereotyping," *Social Cognition*, vol. 34, no. 6, pp. 523–543, 2016.

[24] R. J. Rydell, D. L. Hamilton, and T. Devos, "Now they are American, now they are not: valence as a determinant of the inclusion of African Americans in the American identity," *Social Cognition*, vol. 28, no. 2, pp. 161–179, 2010.

[25] I. V. Blair, J. E. Ma, and A. P. Lenton, "Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery," *Journal of Personality and Social Psychology*, vol. 81, no. 5, pp. 828–841, 2001.

[26] N. Dasgupta and S. Asgari, "Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping," *Journal of Experimental Social Psychology*, vol. 40, no. 5, pp. 642–658, 2004.

[27] S. T. Dunlap and J. M. Barth, "Career stereotypes and identities: implicit beliefs and major choice for college women and men in STEM and female-dominated fields," *Sex Roles*, vol. 81, no. 9-10, pp. 548–560, 2019.

[28] L. R. Ramsey, D. E. Betz, and D. Sekaquaptewa, "The effects of an academic environment intervention on science identification among women in STEM," *Social Psychology of Education*, vol. 16, no. 3, pp. 377–397, 2013.

[29] J. G. Stout, N. Dasgupta, M. Hunsinger, and M. A. Mcmanus, "STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM)," *Journal of Personality and Social Psychology*, vol. 100, no. 2, pp. 255–270, 2011.

[30] A. G. Greenwald, D. E. McGhee, and J. L. K. Schwartz, "Measuring individual differences in implicit cognition: the implicit association test," *Journal of Personality and Social Psychology*, vol. 74, no. 6, pp. 1464–1480, 1998.

[31] P. G. Devine, "Stereotypes and prejudice: their automatic and controlled components," *Journal of Personality and Social Psychology*, vol. 56, no. 1, pp. 5–18, 1989.

[32] B. Wittenbrink, C. M. Judd, and B. Park, "Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures," *Journal of Personality and Social Psychology*, vol. 72, no. 2, pp. 262–274, 1997.

[33] B. Kurdi, T. C. Mann, T. E. S. Charlesworth, and M. R. Banaji, "The relationship between implicit intergroup attitudes and beliefs," *Proceedings of the National Academy of Sciences*, vol. 116, no. 13, pp. 5862–5871, 2019.

[34] A. G. Greenwald, B. A. Nosek, and M. R. Banaji, "Understanding and using the implicit association test: I. An improved scoring algorithm," *Journal of Personality and Social Psychology*, vol. 85, no. 2, pp. 197–216, 2003.

[35] B. A. Nosek, M. R. Banaji, and A. G. Greenwald, "Harvesting implicit group attitudes and beliefs from a demonstration web site," *Group Dynamics: Theory, Research, and Practice*, vol. 6, no. 1, pp. 101–115, 2002.

[36] B. A. Nosek, F. L. Smyth, J. J. Hansen et al., "Pervasiveness and correlates of implicit attitudes and stereotypes," *European Review of Social Psychology*, vol. 18, no. 1, pp. 36–88, 2007.

[37] A. Locksley, E. Borgida, N. Brekke, and C. Hepburn, "Sex stereotypes and social judgment," *Journal of Personality and Social Psychology*, vol. 39, no. 5, pp. 821–831, 1980.

[38] Z. Kunda and P. Thagard, "Forming impressions from stereotypes, traits, and behaviors: a parallel-constraint-satisfaction theory," *Psychological Review*, vol. 103, no. 2, pp. 284–308, 1996.

[39] T. P. Carpenter, R. Pogacar, C. Pullig et al., "Survey-software implicit association tests: a methodological and empirical analysis," *Behavior Research Methods*, vol. 51, no. 5, pp. 2194–2208, 2019.

[40] S. T. Fiske and S. L. Neuberg, "A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation," *Advances in Experimental Social Psychology*, vol. 23, pp. 1–74, 1990.

[41] H. Jeffreys, *The Theory of Probability*, Oxford University Press, Oxford, UK, 1961.

[42] M. D. Lee and E. J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course*, Cambridge University Press, Cambridge, UK, 2014.

[43] M. A. Hogg, *The Social Psychology of Group Cohesiveness: From Attraction to Social Identity*, Harvester Wheatsheaf, London, UK, 1992.

[44] M. A. Hogg, "Group cohesiveness: a critical review and some new directions," *European Review of Social Psychology*, vol. 4, no. 1, pp. 85–111, 1993.

[45] D. Abrams and M. A. Hogg, "Comments on the motivational status of self-esteem in social identity and intergroup discrimination," *European Journal of Social Psychology*, vol. 18, no. 4, pp. 317–334, 1988.

[46] H. Tajfel and J. Turner, "An integrative theory of intergroup conflict," in *The Social Psychology of Intergroup Relations*, W. Austin and S. Worchel, Eds., pp. 33–47, Brooks/Cole Publishing Company, Salt Lake City, UT, USA, 1979.

[47] M. A. Hogg, D. J. Terry, and K. M. White, "A tale of two theories: a critical comparison of identity theory with social identity theory," *Social Psychology Quarterly*, vol. 58, no. 4, p. 255, 1995.

[48] Z. Kunda and S. J. Spencer, "When do stereotypes come to mind and when do they color judgment? a goal-based theoretical framework for stereotype activation and application," *Psychological Bulletin*, vol. 129, no. 4, pp. 522–544, 2003.

[49] M. Snyder and W. B. Swann, "Hypothesis-testing processes in social interaction," *Journal of Personality and Social Psychology*, vol. 36, no. 11, pp. 1202–1212, 1978.

[50] R. S. Nickerson, "Confirmation bias: a ubiquitous phenomenon in many guises," *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[51] J. Fyock and C. Stangor, "The role of memory biases in stereotype maintenance," *British Journal of Social Psychology*, vol. 33, no. 3, pp. 331–343, 1994.

[52] J. M. Darley and P. H. Gross, "A hypothesis-confirming bias in labeling effects," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 20–33, 1983.

[53] R. Weber and J. Crocker, "Cognitive processes in the revision of stereotypic beliefs," *Journal of Personality and Social Psychology*, vol. 45, no. 5, pp. 961–977, 1983.

[54] M. Hewstone, C. N. Macrae, R. Griffiths, A. B. Milne, and R. Brown, "Cognitive models of stereotype change: (5). Measurement, development, and consequences of subtyping," *Journal of Experimental Social Psychology*, vol. 30, no. 6, pp. 505–526, 1994.

[55] M. Rothbart, "Category-exemplar dynamics and stereotype change," *International Journal of Intercultural Relations*, vol. 20, no. 3-4, pp. 305–321, 1996.

[56] M. Hewstone, "Revision and change of stereotypic beliefs: in search of the elusive subtyping model," *European Review of Social Psychology*, vol. 5, no. 1, pp. 69–109, 1994.

[57] C. Badea and D. K. Sherman, "Self-affirmation and prejudice reduction: when and why?" *Current Directions in Psychological Science*, vol. 28, no. 1, pp. 40–46, 2019.

[58] C. G. Lord, M. R. Lepper, and E. Preston, "Considering the opposite: a corrective strategy for social judgment," *Journal of Personality and Social Psychology*, vol. 47, no. 6, pp. 1231–1243, 1984.

[59] C. K. Morewedge, H. Yoon, I. Scopelliti, C. W. Symborski, J. H. Korris, and K. S. Kassam, "Debiasing decisions: improved decision making with a single training intervention," *Policy Insights from the Behavioral and Brain Sciences*, vol. 2, no. 1, pp. 129–140, 2015.

[60] B. Gawronski, "Six lessons for a cogent science of implicit bias and its criticism," *Perspectives on Psychological Science*, vol. 14, no. 4, pp. 574–595, 2019.

[61] H. Blanton, J. Jaccard, E. Strauts, G. Mitchell, and P. E. Tetlock, "Toward a meaningful metric of implicit prejudice," *Journal of Applied Psychology*, vol. 100, no. 5, pp. 1468–1481, 2015.

[62] J. Cone and M. J. Ferguson, "He did what? the role of diagnosticity in revising implicit evaluations," *Journal of Personality and Social Psychology*, vol. 108, no. 1, pp. 37–57, 2015.