Journal of Experimental Social Psychology 79 (2018) 42–50

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Registered Report

The effects of category and physical features on stereotyping and evaluation☆

Debbie S. Ma[a,*], Joshua Correll[b], Bernd Wittenbrink[c]

[a] California State University Northridge, USA
[b] University of Colorado at Boulder, USA
[c] University of Chicago Booth School of Business, USA

ARTICLE INFO

Keywords:
Stereotyping
Prejudice
Prototypicality
Phenotype
Feature-based judgment

ABSTRACT

Stereotyping and prejudice researchers have provided numerous demonstrations that the greater a target's prototypicality, the more similar attitudes and inferences will be to the attitudes and stereotypes perceivers have about the group. However, research to date has yet to also test for a possible quadratic association relating target prototypicality to judgment. The current research offers an extension of existing research by testing for both linear and quadratic relationships between target prototypicality and stereotyping using an implicit measure of stereotyping. In Study 1, we tested for linear and quadratic associations between racial prototypicality and stereotyping of Black and White males, while also manipulating the valence of the stereotypes. Study 2 offered a conceptual replication of Study 1 and tested for linear and quadratic associations between gender prototypicality and stereotyping of White males and White females, while again manipulating the valence of these gender stereotypes. Across both studies we replicated previous research showing a positive, linear effect of prototypicality on stereotyping, such that targets greater in prototypicality elicited greater stereotyping. We also found evidence of a quadratic effect of prototypicality, such that average prototypic targets elicited the most stereotyping. Finally, we observed that negative, rather than positive, stereotypes drove both the linear and quadratic effects we report.

1. The effects of category membership and physical features on stereotyping and evaluation

In Florida, like many states in the U.S., active attempts to mitigate racial discrimination in prison sentencing have resulted in statutes clearly outlining appropriate sentence lengths given legitimate factors like crime severity and previous criminal record. Prior to the development of these statutes, Blacks received harsher sentences than Whites, even after controlling for lawful predictors of sentence length (Bales, 1997). By and large, these statues have been touted as successful. The Florida Department of Corrections, for example, has stated that since implementing these objective sentencing standards and limiting judges' discretion in sentencing, there is no evidence for a measurable effect of race (Bales, 1997). A reanalysis of these data corroborated that a defendant's identification as Black or White did indeed not influence sentencing. However, the defendant's physical appearance did impact their sentence. Individuals – both Black and White – who had more "Afrocentric" features (e.g., darker skin tone, wider nose, thicker lips) received harsher prison sentences (Blair, Judd, & Chapleau, 2004). The effect was nontrivial – when comparing those who were 1 standard deviation below the mean to those who were 1 standard deviation above the mean in perceived Afrocentricity, a 7–8 month difference in sentence length was found – even after controlling for criminal record and race. Similar results have been found examining death penalty-eligible cases in Pennsylvania (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Here, researchers found that, when Black defendants committed a crime against a White individual, the likelihood of a death sentence increased as a function of the defendant's "stereotypical" looks, as judged by a convenience sample of participants from a photograph.

2. Feature effects on prejudice and stereotyping over and above category membership

Laboratory studies confirm this pattern of covariation. Research generally reveals a positive, linear relationship between target

☆ Primary support for this work was provided by National Science Foundation (1226143) to the first author. We would like to thank members of the Stereotyping and Intergroup Prejudice Lab at California State University Northridge, Penny Visser, and Robert Livingston for their helpful comments on the work.
* Corresponding author at: Department of Psychology, California State University, Northridge, 18111 Nordhoff Street, Northridge, CA 91330, USA.
E-mail address: Debbie.Ma@csun.edu (D.S. Ma).

https://doi.org/10.1016/j.jesp.2018.06.008
Received 10 March 2018; Received in revised form 15 June 2018; Accepted 23 June 2018

prototypicality and category-consistent judgments. The greater an individual's prototypicality (be it operationalized by a set of features or just skin tone in the context of judgments of Blacks), the more that person is stereotyped or evaluated in a manner consistent with the group stereotypes and evaluations. The argument is that, if, as a category, Blacks are disliked (or perceived as athletic), then the more prototypic a person is of Blacks, the more disliked (athletic) the individual should seem (e.g., Blair & Judd, 2010; Maddox, 2004). This linear relationship between features and judgment has been demonstrated across a variety of attitude-related domains: evaluation/prejudice, explicit stereotyping, and implicit stereotyping.

In the domain of prejudice, individuals expressed greater negativity toward darker-skinned Blacks than lighter-skinned Blacks and also rated them as less attractive (Maddox & Gray, 2002; see also Hagiwara, Kashy, & Cesario, 2012). Similar skin tone results have been demonstrated in American Hispanics' and Chileans' attitudes toward lighter- and darker-complected Latinos (Uhlmann, Dasgupta, Elgueta, Greenwald, & Swanson, 2002). Convergent evidence using functional magnetic resonance imaging (fMRI) has revealed that exposure to dark-skinned White males (non-prototypic Whites) elicited greater amygdala activation (often taken as an index of threat perception) relative to light-skinned White males (Ronquillo et al., 2007). Finally, Livingston and Brewer (2002) investigated the extent to which Black prototypicality impacted implicit prejudice over and above race. They showed that highly prototypic Black targets elicited more prejudice than less prototypic targets on an implicit measure of prejudice. Across all of these studies, researchers compared targets that were low versus high in prototypicality and showed that targets who were high in prototypicality were judged in a manner more consistent with group attitudes.

In addition to evaluation, research supports a positive, linear link between prototypicality and stereotyping. For example, with regard to explicitly measured stereotypes, Anderson and Cromwell (1977) showed that darker-complected individuals were viewed as less intelligent, consistent with the cultural stereotype that Blacks are not smart. In a more recent study, Maddox and Gray (2002) asked participants to list stereotypic traits that characterize dark- and light-skinned Blacks. Using Devine's (1989) Black stereotype trait list, they coded participants' responses and found that people listed significantly more Black stereotypic traits in response to darker-skinned Blacks than lighter-skinned Black. Conversely, participants listed fewer counter-stereotypic traits when describing dark-skinned versus lighter-skinned Blacks. Among these traits, there was also an effect for the valence of traits participants listed for the two groups of Blacks, such that they listed more positive traits when describing lighter-skinned Black and more negative traits when describing darker-skinned Blacks.

Likewise, Blair, Judd, Sadler, and Jenkins (2002) employed an impression formation paradigm in which participants were presented with a description of an individual that varied in terms of stereotypically Black behavior and valence. Participants were then given photographs of Blacks and Whites who varied in Afrocentricity and were asked to rate the probability that each photograph was the individual being described. Researchers found that more Afrocentric targets were rated as more likely to be the person in the stereotypically Black descriptions (see also Blair, Judd, & Fallman, 2004). Moreover, racial prototypicality has been found to moderate the decision to shoot using a computer video game in which participants are asked to execute shoot/don't shoot decisions in response to armed and unarmed Black and White males. Ma and Correll (2011) replicated previous research showing racial bias in shoot decisions (e.g., more false alarms in response to unarmed Blacks than Whites; Correll, Park, Judd, & Wittenbrink, 2002), and observed that this racial bias increased linearly with the targets' prototypicality. Unlike some of the previously documented research, these studies included targets that varied continuously in terms of prototypicality. However, these studies only report tests for positive, linear associations between prototypicality and judgment (but see the General Discussion). The current research tests another possibility –features also relate to judgment in a curvilinear fashion. In the next section, we explain the rationale for our research.

3. Perspectives on category structure

The idea that categories are graded, that the members of a category vary in the extent to which they fit the category, has been a central topic of research in cognitive psychology. Researchers agree that in many cases different members of a category are not equivalently examples of the category (Rips, Shoben, & Smith, 1973; Rosch, 1973). Although robins, chickens, and flamingoes all belong to the category, *bird*, participants rate these animals differently in terms of the goodness with which they represent the category (Barsalou, 1983). The question of how graded categories are structured and what makes members more or less typical, however, remains up for debate. Two perspectives that are especially relevant to the current discussion are the *family resemblance* and *ideals* perspectives. According to the former model, an exemplar's typicality depends on its similarity to the category's central tendency (Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). A category's most representative exemplar is therefore closest to the category's central tendency. Imagine, for example, Black faces that have been rated in terms of prototypicality. Family resemblance models would suggest that the most representative exemplar is one close to the mean or median of the sample. By contrast, the category ideal perspective suggests that the most representative exemplar exists at the periphery of the category: it is the extreme, rather than the average. For a visual depiction of the family resemblance and ideal perspectives, see Fig. 1. The ideal view is consistent with the idea that more extreme faces are better representations of the category and thus elicit stronger activation of category judgments (this is consistent with the linear effects observed in race-related work). But the family-resemblance

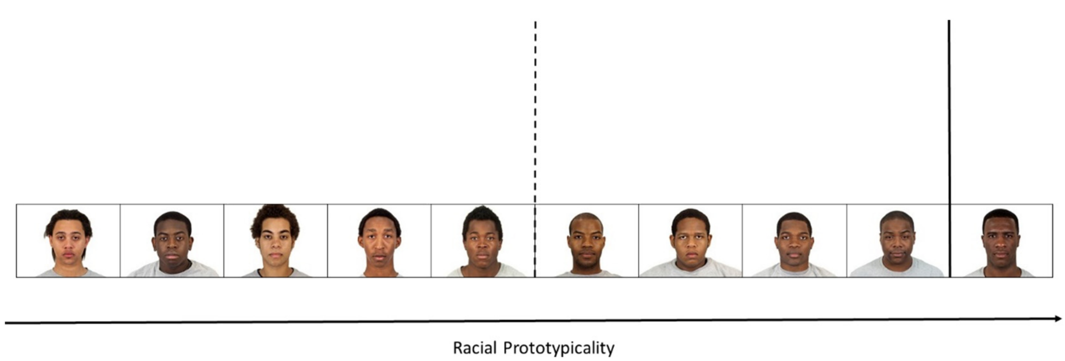

Fig. 1. Visual depiction of the family resemblance and ideal perspectives. Faces assembled in order of least to most racially prototypic (left to right). The dotted line represents a theoretical average. Faces within the average range are prototypic, as defined by the family resemblance perspective. The solid line represents a theoretical category ideal. Faces toward the right of the distribution are more prototypic, as defined by the ideal.

perspective argues that the best exemplar is the central tendency. This perspective therefore suggests that average targets should trigger greater category-consistent judgments. Exemplars that are low on Afrocentricity should evoke the category less, but (perhaps counter-intuitively) exemplars that are extremely high on Afrocentricity should also evoke the category less. This hypothesis would be borne out by a quadratic relationship between features and judgments, as opposed to the linear relationship between prototypicality and category judgments that others have reported. One of the goals of the current research is to explore the possibility that features and judgment may be related in a nonlinear fashion.

4. The current research

The goal of the current research is to test for a possible quadratic relationship between features and judgment in an experimental context. Whereas all of the existing social psychological research has documented a positive, linear relationship between features and stereotyping or features and evaluation, this research has not considered a non-linear relationship, which would require including more than two levels of prototypicality (however, see Dunham, Dotsch, Clark, & Stepanova, 2016; Dunham, Stepanova, Dotsch, & Todorov, 2014; Stepanova & Strube, 2012 for research on racial categorization). Across two studies, we utilize an implicit measure of stereotyping and test for linear and quadratic relationship between prototypicality and judgment among Black and White males (Study 1) and White males and females (Study 2).

5. Study 1

Study 1 assessed the influence of features on stereotyping of Black and White males using a Lexical Decision Task (LDT, Neely, 1977), a commonly used implicit measure of stereotyping. We examined the degree to which prototypicality among Black and White males activated Black and White stereotypes, which we also varied by valence.

6. Method

6.1. Participants and research design

Eighty-one White students (59 female, 21 male, 1 did not report) at a large west coast university participated in this experiment in exchange for partial course credit. No participants were excluded from the analysis. We conducted a sensitivity power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) with an alpha of 0.05, two-tailed test and powered at 80%. This analysis indicated that we had the power to detect an effect size of $d = 0.315$. The data were collected in accordance with the university IRB protocol. Participants were 19.84 (SD = 2.96) years old on average. The study involved a 2 (Prime Race: Black and White) × prototypicality (continuously measured) × 2 (Word Type: Black stereotype and White stereotype) × 2 (Word Valence: Negative and Positive) repeated measures design. We report all of the measured variables herein.

6.2. Stimuli

6.2.1. Word selection

Black and White stereotypic words were obtained after extensive pre-testing to equate for valence, frequency, and length. Specifically, 90 stereotypic Black and White traits were accumulated from a variety of published resources (Devine, 1989; Devine & Elliot, 1995; Dovidio, Evans, & Tyler, 1986; Wittenbrink, Judd, & Park, 1997). A convenience sample of 82 students recruited from a Midwestern university rated these traits for valence (1 = extremely negative, 7 = extremely positive) or the extent to which each word was part the cultural stereotype of Black American males, White American males, Black American females, and White American females (1 = definitely not a part of the cultural stereotype, 7 = definitely a part of the cultural stereotype). Each participant received only 1 of the 5 surveys. We selected the following sample of stereotypes based on these ratings: Black-positive (*athlete*, *musical*, *streetwise*, *social*, *soulful*), Black-negative (*danger*, *lazy*, *poverty*, *rude*, *stupid*); White-positive (*ambitious*, *rich*, *smart*, *peaceful*, *wealthy*) and White-negative (*boring*, *greedy*, *materialistic*, *selfish*, *stuffy*). Pronounceable, non-word letter strings of equal length were also used in the task for the non-word trials.

6.2.2. Target photographs

Digital photographs of 30 Black and White male targets were selected from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015). The database includes a host of subjective and objective ratings. Across race, targets were equated for attractiveness and masculinity.

6.2.3. Racial prototypicality

Target ratings of racial prototypicality are available as part of the Chicago Face Database norming data. These judgments were obtained from a sample who rated each of the faces on a 0–100 semantic differential scale (0 = Very White/Eurocentric; 100 – Very Black/Afrocentric). The decision to measure racial prototypicality using this scale was motivated by prior research indicating that Black and White prototypicality are inversely predicted by the same set of facial features (Blair & Judd, 2010). Based on these ratings, we selected five Black and five White targets at low, average, and high racial prototypicality. In the subsequent analyses, we subtracted the ratings from 100 for the White targets' so that the scores for both Black and White targets were scored such that higher values corresponded with greater racial prototypicality. Prototypicality ratings were ($M = 75.97$; $SD = 1.79$) for low prototypic targets, ($M = 80.01$; $SD = 1.33$) for average prototypic targets, and ($M = 85.07$; $SD = 1.56$) for high prototypic targets, though we treated the data as continuous in the analysis.

6.3. Procedure

6.3.1. Lexical Decision Task (LDT)

After providing verbal consent, participants were seated at a PC where they read instructions for the LDT and completed the remainder of the study. The LDT began with 10 test trials to familiarize participants with making word/non-word judgments. Participants were told to press a key labeled “non-word” if the letter string they saw was not an actual word and were told to press a separate key labeled “word” if the letter string they saw was a word. Primes and words used for these practice trials were unrelated to race categories. After the practice trials, participants were given an opportunity to ask the experimenter questions about the task before completing the experimental portion of the study. Every test trial began with a fixation cross that was displayed for 500 ms, followed by a randomly selected face prime for 250 ms, a blank screen for 135 ms, and finally a letter string, which remained on the screen until participants classified the string as a word or non-word using designated keys on the keyboard. Participants completed 360 test trials in total. Half of the trials included non-word targets and half were the critical trials on which a word was presented. Critical trials crossed prime race, prototypicality, word type, and word valence.

6.3.2. Demographics

Upon completing the LDT, the participants provided basic demographic information and received a thorough debriefing. Once all participants completed the experiment, they were thanked and excused.

6.4. Analytic approach

We trimmed the data of trials on which participants responded faster than 100 ms or slower than 4000 ms and where participants responded incorrectly (3.56% of trials). We maintained a wide response

Table 1
Results from Study 1 analyses.

Predictor	Mean	n	SD	t-value	*p*-value
Prime race	−0.0127	81	0.0534	−2.13	0.04
Linear prototypicality	−0.0005	81	0.007	−0.58	0.56
Word type	−0.0081	81	0.0413	−1.77	0.08
Word valence	−0.0153	81	0.0436	−3.16	0.002
Prime race × linear prototypicality	0.0001	81	0.0082	0.12	0.9
Prime race × word type	0.0199	81	0.0432	4.14	< 0.001
Prime race × word valence	0.005	81	0.0465	0.96	0.34
Linear prototypicality × word type	−0.0014	81	0.0077	−1.69	0.10
Linear prototypicality × word valence	−0.0012	81	0.0096	−1.15	0.25
Word type × word valence	−0.0283	81	0.0406	−6.29	< 0.001
Prime race × linear prototypicality × word type	0.0008	81	0.0076	0.93	0.35
Prime race × linear prototypicality × word valence	−0.0006	81	0.0086	−0.65	0.52
Prime race × word type × word valence	−0.0144	81	0.04	−3.25	0.002
Linear prototypicality × word type × word valence	0.0017	81	0.0067	2.33	0.02
Prime race × linear prototypicality × word type × word valence	−0.0016	81	0.0066	−2.27	0.03
Quadratic prototypicality	0.0002	81	0.0024	0.85	0.40
Prime race × quadratic prototypicality	0	81	0.002	0.18	0.85
Quadratic prototypicality × word type	0.0006	81	0.0023	2.41	0.02
Quadratic prototypicality × word valence	0.0003	81	0.0022	1.14	0.26
Prime race × quadratic prototypicality × word type	−0.0009	81	0.0021	−3.94	< 0.001
Prime race × quadratic prototypicality × word valence	−0.0001	81	0.0023	−0.55	0.58
Quadratic prototypicality × word type × word valence	−0.0003	81	0.0019	−1.3	0.20
Prime race × quadratic prototypicality × word type × word valence	0.0005	81	0.002	2.38	0.02

window to preserve as much of the data as possible and note that adjusting these cutoff points did not affect the significance levels of the key results we report. Reaction times were log-transformed and submitted to multilevel modeling treating participants as a random factor. We ran separate multiple regression analyses for each participant, then (across all participants) computed average slopes for each predictor in the model and compared those slopes to zero using a one-sample *t*-test.

7. Results

We regressed log-transformed reactions times on prime race (contrast coded: Black = +1; White = −1), linear prototypicality[1] (i.e., mean-centered racial prototypicality), quadratic prototypicality (i.e., mean-centered racial prototypicality squared), word type (contrast coded: Black = −1; White = +1), word valence (contrast coded: negative = −1; positive = +1), and all higher-order interactions. Here, we focus on the theoretically-meaningful effects, but present the full set of results in Table 1.

7.1. Race-based judgments

First, we explore interactions of prime race with (a) word type and (b) word valence, which can be viewed as tests of category-based stereotype activation and prejudice, respectively. We observed a significant effect of racial stereotype activation, as indicated by a prime race × word type interaction, $t(80) = 3.69$, $p < .001$, $d = 0.41$. Following Black primes, participants were marginally faster to recognize Black stereotypic words ($M = 6.53$, $SD = 0.19$) relative to White stereotypic words ($M = 6.55$, $SD = 0.22$), $t(80) = 1.99$, $p = .05$, $d = 0.22$. By contrast, White primes led participants to respond faster to White ($M = 6.53$, $SD = 0.20$) compared to Black stereotypic words ($M = 6.58$, $SD = 0.21$), $t(80) = -3.79$, $p < .001$, $d = 0.42$. There was no evidence for racial prejudice – the prime race × word valence interaction was not significant, $t(80) = 0.26$, $p = .80$, $d = 0.03$. However, the prime race × word type × word valence interaction was significant, $t(80) = -2.59$, $p = .01$, $d = 0.29$. The pattern of this interaction suggests that stereotype activation (the prime race × word type interaction) was greater when the target word was negative in valence. On negative target trials, the prime race × word type interaction was significant ($M = 0.027$, $SD = 0.054$), $t(80) = 4.47$, $p < .001$, $d = 0.50$, but this effect was not significant on positive target trials ($M = 0.005$, $SD = 0.054$), $t(80) = 0.92$, $p = .36$, $d = 0.10$.

7.2. Linear effects of prototypicality

Next, we examine the degree to which effects of prime race (the Black vs. White difference) vary as a function of the racial prototypicality of the prime. We test whether the effect of race *linearly* related to prototypicality (e.g., is the Black-White difference stronger among high-prototypicality faces than among low-prototypicality faces). As we consider the linear relationship, we will again draw distinctions between race-based stereotype activation and race-based prejudice, which will each be tested in turn. First, the prime race × linear prototypicality × word type interaction tested whether racial stereotype activation was moderated by prototypicality in a linear form. We found no evidence of this effect, $t(80) = -0.64$, $p = .53$, $d = 0.07$. We also did not find any evidence that racial prejudice was linearly related to prototypicality – the prime race × linear prototypicality × word valence interaction was not significant, $t(80) = -0.64$, $p = .52$, $d = 0.07$. However, the prime race × linear prototypicality × word type × word valence interaction was significant, $t(80) = -2.37$, $p = .02$, $d = 0.26$. Similar to the effects described in the previous section, the pattern suggests that effects of stereotype activation are stronger on negatively valenced target words. To decompose this interaction, we therefore examined the linear effect of features on stereotype activation separately for negative and positive trials. On negative target trials, the prime race × linear prototypicality × word type was significant, $t(80) = 2.53$, $p = .01$, $d = 0.28$. For negative target words, as prototypicality increased, stereotype activation increased. The prime race × word type interaction was not significant at low prototypicality ($M = -0.004$, $SD = 0.049$), $t(80) = -0.70$, $p = .48$, $d = 0.08$, but was at both average ($M = 0.027$, $SD = 0.054$), $t(80) = 4.47$, $p < .001$, $d = 0.50$ and high prototypicality ($M = 0.015$, $SD = 0.048$), $t(80) = 2.93$, $p = .004$, $d = 0.33$, respectively. The same effect on positive stereotype trials was not significant, $t(80) = -0.64$, $p = .53$, $d = 0.07$. For the sake of

[1] In a follow-up analysis, we tested the simultaneous effects of prototypicality and luminance on stereotyping. Hagiwara et al. (2012), for example, observed independent effects of luminance and physical features on attitude. Critically, inclusion of luminance and higher-order interactions in the current model did not alter the significance of the key prototypicality effects we report herein.

completeness, we also decomposed the interaction by prime race. On Black prime trials, the linear prototypicality × word type × word valence interaction was not significant, $t(80) = -0.36$, $p = .72$, $d = 0.04$. On White prime trials, however, the linear prototypicality × word type × word valence interaction was significant, $t(80) = 2.76$, $p = .007$, $d = 0.31$.

7.3. Quadratic effects of prototypicality

Finally, we examined the moderating role of prototypicality in a quadratic form: is the effect of race stronger among average faces (i.e., medium prototypicality) than among extreme faces? First, we tested whether the quadratic form of prototypicality moderated racial stereotyping (e.g., was stereotype activation more pronounced among average faces) and indeed, the prime race × quadratic prototypicality × word type interaction was significant, $t(80) = -3.66$, $p < .001$, $d = 0.41$. At low ($M = 0.000$, $SD = 0.033$) and high ($M = 0.007$, $SD = 0.040$) racial prototypicality, there was no evidence of a prime race × word type interaction, $t(80) = 0.77$, $p = .44$, $d = 0.09$, and $t(80) = 1.63$, $p = .11$, $d = 0.18$, respectively. However, the prime race × word type interactions at average prototypicality ($M = 0.0012$, $SD = 0.041$), $t(80) = 3.69$, $p < .001$, $d = 0.41$. Next, we tested whether the quadratic form of prototypicality moderated racial prejudice to determine whether the effect of prejudice was more pronounced among average faces. There was no evidence of such an effect, as revealed by a non-significant prime race × quadratic prototypicality × word valence interaction, $t(80) = 0.16$, $p = .87$, $d = 0.02$. Lastly, we wanted to determine whether the moderating effect of the quadratic form of prototypicality on racial stereotype activation was further moderated by the valence of the stereotypes. This was tested by the 4-way interaction between prime race × quadratic prototypicality × word type × word valence interaction, which was significant, $t(80) = 2.55$, $p = .01$, $d = 0.28$. On negative stereotype trials, there was a negative quadratic effect of racial prototypicality on stereotype activation. The prime race × quadratic prototypicality × word type interaction was significant, $t(80) = -4.29$, $p < .001$, $d = 0.48$. The estimated means of the prime race × word type interaction for negative stereotypes at low, average, and high prototypicality, which we reported above, suggest that racial bias was most pronounced by targets that were average in terms of racial prototypicality. On positive stereotype trials, by contrast, the prime race × quadratic prototypicality × word type interaction was not significant, $t(80) = -0.88$, $p = .38$, $d = 0.10$.

8. Discussion

Study 1 offered initial support that the relationship between prototypicality and categorical judgments can be quadratic, as well as linear. Both the linear and quadratic relationship between prototypicality and stereotype activation were limited or more pronounced among negative stereotypes. On negative stereotype trials, more prototypic targets elicited greater stereotype activation, but stereotype activation was most pronounced for the average prototypic targets. This finding comports with prior research in which the effect of features on stereotypes appeared to be stronger for negative than positive stereotypes (Blair, 2006; Blair et al., 2002; Blair, Chapleau, & Judd, 2005; Blair, Judd, & Fallman, 2004). Because our research, and that of much of the published research, is limited to stereotypes about Black versus White targets, Study 2 sought to re-test the hypothesis that prototypicality might have a curvilinear relationship with category-based judgment in a different context.

9. Study 2

The goal of Study 2 was to conceptually replicate the findings reported in Study 1 using a different social category. This step is important, because it will allow us to test whether the findings of Study 1 are specific to Black and White males or whether they are more generalizable. As we review in the Introduction, much of the research in this space has been limited to investigations involving Black men and White men and much less research has been done examining gender prototypicality and stereotyping (for an exception see Kahn, Unzueta, Davies, Alston, & Lee, 2015). Study 2 also represents the first study to examine how gender prototypicality influences judgments in the context of an implicit stereotyping task using visual primes.

10. Method

10.1. Participants and research design

Participants were 116 students (84 female, 28 male, 4 did not report) at a large west coast university in exchange for partial course credit. The sample was racially diverse (54 Latino, 36 White, 7 Asian, 7 Black, 4 biracial/multiracial, 1 Native American, 3 Other, and 4 did not report). A priori power analysis computed with G*Power established power at > 0.85. The data were collected in accordance with the university IRB protocol. Participants were 19.60 (SD = 3.99) years old on average. The study involved a 2 (Prime Gender: male and female) × 3 (Prototypicality: continuously measured) × 2 (Word Type: male stereotype and female stereotype) × 2 (Word Valence: negative and positive) repeated measures design. The sample size was determined by the number of participants that we were able to gather in one academic semester. No participants were dropped from the analysis and we report all of the data that were collected.

10.2. Stimuli

10.2.1. Word selection

Males and female stereotypic traits were accumulated from a variety of sources (i.e., Cejka & Eagly, 1999; Diekman & Eagly, 2000; Eagly & Sczesny, 2009; Prentice & Carranza, 2002) and submitted to a pre-test in order to identify words that were stereotypic of White American males or females and clearly positively or negatively valenced. Participants were shown words in random order and asked to rate either the valence of each word (1 = extremely negative, 7 = extremely positive) or the extent to which each word was part the cultural stereotype of White American males and White American females (1 = definitely not a part of the cultural stereotype, 7 = definitely a part of the cultural stereotype). This produced 3 different surveys. A convenience sample of 76 students recruited from a Midwestern university completed 1 of the 3 surveys. From these ratings, we selected 36 words (9 for each gender-by-valence category): female-positive (*affectionate*, *cautious*, *gentle*, *kind*, *motherly*, *polite*, *sensitive*, *social*, *trendy*), female-negative (*fussy*, *bitchy*, *catty*, *cliquish*, *moody*, *nagging*, *naïve*, *picky*, *whiny*), male-positive (*athletic*, *daring*, *forthright*, *masculine*, *powerful*, *resolute*, *strong*, *streetwise*, *tough*), and male-negative (*coarse*, *crude*, *forceful*, *messy*, *reckless*, *rude*, *sloppy*, *stingy*, *violent*). As in Study 1, words were equated for frequency, length, and valence (i.e., positive words were equally positive and negative words were equally negative). Pronounceable, non-words of equal length were also used.

10.2.2. Target photographs

Digital photographs of 30 White male and White female targets were selected from the Chicago Face Database (Ma et al., 2015). We selected 5 males and 5 females who were rated as low, average, and high in terms of gender prototypicality. Targets were equated for attractiveness.

10.2.3. Gender prototypicality

We used ratings of masculinity and femininity from the Chicago Face Database to select five male and five female targets who were judged to be low, average, and high in gender prototypicality. For males, masculinity ratings served as the measure of gender prototypicality, whereas for females, we used femininity ratings. Masculinity and femininity were measured on a 7-point Likert scale (1 = Not at All; 7 = Extremely). Prototypicality ratings were ($M = 4.16$; $SD = 0.19$) for low prototypic targets, ($M = 4.54$; $SD = 0.30$) for average prototypic targets, and ($M = 4.95$; $SD = 0.26$) for high prototypic targets, though we treated the data as continuous in the analysis.

10.3. Procedure

10.3.1. Lexical Decision Task (LDT)

The procedures followed those of Study 1. After providing verbal consent, participants read the instructions for the LDT on the computer. Participants completed 10 practice trials, followed by 432 test trials. The task parameters were otherwise identical to Study 1. Half of the trials featured non-words. The other 216 trials included 27 trials of each prime gender, word type, word valence combinations.

10.3.2. Demographics

Upon completing the LDT, participants provided basic demographic information and received a thorough debriefing and were thanked for their participation.

10.4. Analytic approach

As in Study 1, we trimmed the data of trials on which participants responded faster than 100 ms or slower than 4000 ms and where participants responded incorrectly (4.18% of trials). Adjustments to the cut-off criteria did not influence the significance of the results. Reaction times were log-transformed and submitted to a series of multilevel modeling to test for category-based, linear, and quadratic effects of gender prototypicality. Again, this type of analysis enables us to maximize power in our study.

11. Results

We regressed reaction times on prime gender (contrast-coded: females = −1; males = +1), linear prototypicality (i.e., mean-centered gender prototypicality), quadratic prototypicality (i.e., mean-centered gender prototypicality squared), word type (contrast-coded: male = −1; female = +1), word valence (contrast-coded: negative = −1, positive = +1), and all higher-order interactions. Again, we highlight the critical results here, but present the full table of results in Table 2.

11.1. Gender-based judgments

Our first analyses focus on the interactions of prime gender with (a) word type and (b) word valence, which again reflect category-based stereotype activation and prejudice. We observed the predicted prime gender × word type interaction, $t(115) = 5.74$, $p < .001$, $d = 0.53$, which was indicative of gender stereotype activation. On male prime trials, there was a simple effect of word type, $t(115) = 5.88$, $p < .001$, $d = 0.55$. Male targets following male primes ($M = 6.41$, $SD = 0.17$) were recognized as words significantly faster than female targets following male primes ($M = 6.46$, $SD = 0.17$). On female prime trials, there was no evidence for a simple effect of word type, $t(115) = -1.53$, $p = .13$, $d = 0.14$. Participants were no faster to recognize female targets ($M = 6.42$, $SD = -0.17$) following female primes than male targets ($M = 6.44$, $SD = 0.17$). We also observed a prime gender × word valence, $t(115) = -5.12$, $p < .001$, $d = 0.47$, which indicated that participants preferred males to females. On male prime trials, there was a simple effect of word valence, $t(115) = -6.47$, $p < .001$, $d = 0.60$. Participants responded faster to positive words ($M = 6.41$, $SD = 0.16$) following a male prime than negative words ($M = 6.47$, $SD = 0.17$). On female prime trials, there was no evidence of word valence, $t(115) = 0.57$, $p = .57$, $d = 0.05$. Participants were no faster to recognize negative words ($M = 6.43$, $SD = 0.17$) compared to positive words ($M = 6.42$, $SD = 0.17$) following female primes. Finally, we found a prime gender × word type × word valence interaction, $t(115) = -4.94$, $p < .001$, $d = 0.46$. On negative trials, we observed significant stereotype activation effect with a prime gender × word type interaction, $t(115) = 8.14$, $p < .001$, $d = 0.76$. On positive trials, the prime gender × word type interaction was not significant, $t(115) = -0.06$, $p = .95$, $d = 0.00$.

11.2. Linear effects of prototypicality

Our next set of analyses tested for linear effects of gender prototypicality over and above gender effects. Again, we draw the distinction between gender-based stereotype activation (i.e., interactions between word type) and gender-based prejudice (i.e., interaction between word valence). First, the prime gender × linear prototypicality × word type interaction tests whether gender prototypicality in the linear form moderates gender-based stereotype activation. This interaction asks whether the effect of gender stereotype activation increases as gender prototypicality increases. This effect was significant, $t(115) = 3.21$, $p = .002$, $d = 0.30$. At low prototypicality, there was no evidence of a prime gender × word type interaction ($M = -0.002$, $SD = 0.035$), $t(115) = -0.49$, $p = .62$, $d = 0.05$. However, this effect was significant at both average prototypicality ($M = 0.016$, $SD = 0.030$), $t(115) = 5.74$, $p < .001$, $d = 0.53$, and high prototypicality ($M = 0.012$, $SD = 0.035$), $t(115) = 3.68$, $p < .001$, $d = 0.34$. Next, we examined the prime gender × linear prototypicality × word valence interaction to test whether gender-based prejudice changes linearly with gender prototypicality. This was indeed the case, $t(115) = -2.26$, $p = .03$, $d = 0.21$. At low prototypicality, there was no evidence of a prime gender × word valence interaction ($M = -0.001$, $SD = 0.042$), $t(115) = -0.34$, $p = .73$, $d = 0.03$; however this interaction was significant at average prototypicality ($M = -0.016$, $SD = 0.035$), $t(115) = -5.12$, $p < .001$, $d = 0.48$, and high gender prototypicality ($M = -0.013$, $SD = 0.034$), $t(115) = -4.09$, $p < .001$, $d = 0.38$. Finally, the four-way interaction between prime gender × linear prototypicality × word type × word valence interaction was significant, $t(115) = 2.21$, $p = .03$, $d = 0.21$. On negative stereotype trials, the prime gender × linear prototypicality × word type interaction was significant, $t(115) = 3.94$, $p < .001$, $d = 0.37$. At low prototypicality ($M = 0.001$, $SD = 0.047$), the prime gender × word type interaction was not significant, $t(115) = 0.30$, $p = .76$, $d = 0.03$. However, at average ($M = 0.032$, $SD = 0.043$) and high prototypicality ($M = 0.025$, $SD = 0.047$), the prime gender × word type interactions were significant, $t(115) = 8.14$, $p < .001$, $d = 0.76$ and $t(115) = 5.65$, $p < .001$, $d = 0.52$, respectively. By contrast, on positive stereotype trials, the prime gender × linear prototypicality × word type interaction was not significant, $t(115) = 0.94$, $p = .35$, $d = 0.09$.

11.3. Quadratic effects of prototypicality

Lastly, we examined the moderating role of prototypicality in a quadratic form, which asks whether the effects of gender are stronger among average faces than low or high prototypic faces. The prime gender × quadratic prototypicality × word type interaction tested whether the quadratic form of prototypicality moderated gender stereotype activation (e.g., was stereotype activation more pronounced among average faces) and this effect was significant, $t(115) = -5.27$, $p < .001$, $d = 0.49$. The estimated means of the prime gender × word type interaction (which we reported above) at low, average, and high prototypicality suggested that the targets who were average in terms of gender prototypicality elicited the greatest gender stereotype activation. There was also evidence of a quadratic moderating effect of gender prototypicality on gender-based prejudice, as revealed by a significant prime gender × quadratic prototypicality × word valence interaction, t

Table 2
Results from Study 2 analysis.

Predictor	Mean	n	SD	t-value	*p*-value
Prime gender	0.0036	116	0.0327	1.2	0.23
Linear prototypicality	0.0093	116	0.068	1.48	0.14
Word type	0.0092	116	0.0363	2.74	0.007
Word valence	−0.0137	116	0.0367	−4.03	< 0.001
Prime gender × linear prototypicality	0.0097	116	0.0635	1.65	0.10
Prime gender × word type	0.0161	116	0.0302	5.74	< 0.001
Prime gender × word valence	−0.0164	116	0.0345	−5.12	< 0.001
Linear prototypicality × word type	0.0189	116	0.0675	3.01	0.003
Linear prototypicality × word valence	−0.01	116	0.0711	−1.52	0.13
Word type × word valence	−0.0201	116	0.032	−6.77	< 0.001
Prime gender × linear prototypicality × word type	0.0177	116	0.0596	3.21	0.002
Prime gender × linear prototypicality × word valence	−0.0155	116	0.0738	−2.26	0.03
Prime gender × word type × word valence	−0.0163	116	0.0356	−4.94	< 0.001
Linear prototypicality × word type × word valence	−0.0035	116	0.0744	−0.51	0.61
Prime gender × linear prototypicality × word type × word valence	−0.013	116	0.0635	−2.21	0.03
Quadratic prototypicality	−0.0327	116	0.1758	−2	0.05
Prime gender × quadratic prototypicality	−0.0048	116	0.1706	−0.3	0.76
Quadratic prototypicality × word type	−0.0747	116	0.1764	−4.56	< 0.001
Quadratic prototypicality × word valence	0.0471	116	0.1644	3.09	0.003
Prime gender × quadratic prototypicality × word type	−0.0768	116	0.1569	−5.27	< 0.001
Prime gender × quadratic prototypicality × word valence	0.0646	116	0.1752	3.97	< 0.001
Quadratic prototypicality × word type × word valence	0.0567	116	0.1809	3.37	0.001
Prime gender × quadratic prototypicality × word type × word valence	0.0595	116	0.1746	3.67	< 0.001

(115) = 3.97, $p < .001$, $d = 0.37$. Participants displayed the greatest pro-male bias toward average male and female targets. Again, these means are presented above. Finally, we observed a significant prime gender × quadratic prototypicality × word type × word valence interaction, $t(115) = 3.67$, $p < .001$, $d = 0.34$. On negative stereotype trials, the prime gender × quadratic prototypicality × word type interaction was significant, $t(115) = -6.53$, < 0.001, $d = 0.61$. On positive stereotype trials, however, there was no evidence of a prime gender × linear prototypicality × word type interaction on positive stereotype trials, $t(115) = -0.76$, $p = .45$, $d = 0.07$. We also decomposed the interaction by prime type. When primes were male, there was a significant quadratic prototypicality × word type × word valence interaction, $t(115) = 4.11$, $p < .001$, $d = 0.38$; however, this was not the case for female prime trials, $t(115) = -0.16$, $p = .87$, $d = 0.01$.

12. Discussion

Study 2 extended on the previous research by examining feature-based stereotype activation and feature-based prejudice in a novel domain. In particular, we investigated whether gender prototypicality moderated stereotype activation and prejudice over and above the effect of gender. These data were largely consistent with the findings observed in Study 1, and showed that there are both linear and quadratic effects of prototypicality on stereotype activation and evaluation. Although we found evidence of a linear relationship, the linear trend of these effects significantly reversed at high prototypicality.

13. General discussion

Across two studies and two different social categories, we tested the relationship between category prototypicality and stereotype activation. Recent social psychological research has established a linear relationship between Black racial prototypicality and stereotyping (e.g., Blair, Judd, & Chapleau, 2004; Blair, Judd, & Fallman, 2004; Livingston & Brewer, 2002; Ma & Correll, 2011; Ronquillo et al., 2007); however, we argued that this relationship could also take on a quadratic form. The reported pattern of quadratic effects we observed across both studies uniquely contributes to our understanding of how features relate to judgment. Specifically, previous research either could not or did not include curvilinear terms in their statistical models. Maddox and Gray (2002) operationalized Black prototypicality with a dichotomous variable: dark-skinned Blacks versus light-skinned Blacks (see also Ronquillo et al., 2007). Similarly, Livingston and Brewer (2002) included comparisons between Whites, low prototypic Blacks, and high prototypic Blacks and Hagiwara et al. (2012) likewise cross two levels of skin tone and two levels of racial prototypicality. In order to test for quadratic effects, it would be necessary to include at least a third category involving average skinned/prototypic Blacks, or allow skin tone or prototypicality to vary continuously. In our studies, we had a continuous measure of prototypicality and empirically tested for quadratic effects in our studies. Ultimately, our results suggest that there may be more to consider when we think about how features inform the stereotypes people form and the attitudes they have than what is currently suggested by the existing research on this topic.

The social psychological research suggesting a positive linear relationship linking prototypicality and group judgment ostensibly reveals a divergence from the cognitive psychological literature on category structure that we review in the Introduction. In fact, the current studies found evidence for both linear and quadratic effects. We speculate about several possibilities that might explain this seeming inconsistency and postulate that the observed quadratic effects may not be at odds with the previous findings from social psychology, but rather result from methodological differences between our research and previous findings. First, and most obviously, a critical contribution of the current studies is that we included an operationalization of prototypicality that allowed for a test of the quadratic effect. That said, the notion that the relationship between prototypicality and group judgment be anything by linear and positive represents a departure from the field's current view. One reason why previous research may have been quick to settle on a positive linear relationship between prototypicality and judgment may stem from stimulus set effects. In studies where participants are only presented with low and high prototypic targets, judgments may be characterized by contrast effects, which would produce in a linear association between stereotypic judgment and prototypicality. Researchers have previously established that participants do not typically use all the knowledge that they have to render judgments, but instead utilize subsets of relevant information given the judgment context (e.g., Bodenhausen & Wyer, 1987; Higgins, 1989; Ma, Correll, & Wittenbrink, 2016). As a result, presenting participants with only two types of targets may promote greater perceived differences between them (Schwarz & Bless, 1992). Besides this, it is also worth considering where researchers sampled targets from the prototypicality continuum. In our studies, low, average, and high prototypic face primes were selected at roughly equivalent intervals across ratings of prototypicality. In previous studies, the determination that a target was highly prototypic merely required that it be more prototypic than the low prototypic target. Therefore, it is possible that the high prototypic targets were actually average in prototypicality, which may have produced a more pronounced linear effect.

Although our analytic strategy allowed for a proximal test of prejudice by collapsing across word type and examining responses to positive versus negative words, we acknowledge that our studies were specifically aimed at investigating stereotype activation and are not direct tests of prejudice. First, our selection of the LDT as the measurement task specifically allowed us to assess stereotype activation rather than prejudice. A focus on prejudice might have instead led us to utilize an evaluative priming measure (Wittenbrink & Schwarz, 2007). Second, we selected words that were all related to group stereotypes, rather than general positive and negative words (e.g., sunshine, vacation, vomit, cockroach, etc.). In this way, our studies afforded a clearer test of stereotype activation than prejudice.

Relatedly, although these studies do not provide a direct test of prejudice, we note that negative stereotypes appear to drive the reported results. This may reflect a concept that has been previously documented in the literature referred to as *stereotypic prejudice* – facilitation to negatively valenced stereotypes (e.g., Sassenberg & Wieber, 2005; Wittenbrink, Judd, & Park, 2001). Although we did not anticipate this pattern of results, researchers have long held that negative stereotypes are easier to acquire through both motivational (Brigham, 1971; Hamilton, 1976) and cognitive processes (Hamilton & Gifford, 1976; Spiers, Love, Le Pelley, Gibb, & Murphy, 2017). This may explain why participants in the current studies showed stronger associations between negative stereotypes and their respective groups.

One thing we wish to highlight is the continued need for including more diverse stimuli and samples in social psychological research. For decades, researchers reasoned that examining a basic psychological process in one social context, which in the case of stereotyping and prejudice research has been largely limited to comparisons of Blacks and Whites, was sufficient for theory building (for exceptions see Niemann, Jennings, Rozelle, Baxter, & Sullivan, 1994; Shapiro, Mistler, & Neuberg, 2010). The value of considering other social categories has generally been to reveal possible moderators. This operating standard has served the field well in establishing basic processes, but we may not always be able to even detect moderators and boundary conditions with such a narrow focus. Although the current studies reveal a consistent pattern across Black and White men and White men and women, we cannot be sure these effects generalize to other groups. For instance, as we allude to above, confounds between Black and White stereotypes and low-level features (like luminance) associated with Black and White prototypicality may make it difficult to disentangle feature-based stereotyping from threat associations. Similar idiosyncrasies between features and stereotypes might exist for other groups and in the absence of more inclusive empirical research, we should exercise caution in drawing strong generalizations to other groups.

We also believe that greater diversity is also needed among our participants (Henrich, Heine, & Norenzayan, 2010). Although Study 2 included a diverse sample, Study 1 only recruited White participants, which is consistent with the tradition of most research on stereotyping and prejudice contrasting Blacks and Whites. Indeed, we are only aware of two studies in which non-White participants were investigated in feature-based judgment research. This study involved Latino participants ratings Latino targets and showed that participants had greater prejudice toward darker-complected relative to lighter-complected Latinos (Uhlmann et al., 2002). The second study, was a follow-up that we discussed in Ma and Correll (2011) in which we recruited Black participants to test for a possible outgroup homogeneity explanation. The inclusion of diverse samples is important, because some of the previous research has shown that feature-based stereotyping and prejudice effects are driven primarily by sensitivity to variation in White but not Black features (Ma & Correll, 2011; Ronquillo et al., 2007). Feature variation in White targets, but not Black targets in these studies were responsible for driving the overall effects of features on judgment. This was also the case for Study 1. Several effects of prototypicality (both linear and quadratic) were significant when we looked at variation among White primes, but not when considering Black primes. This suggests that, as with previous studies, participants responded differently to White targets depending on their racial prototypicality, but responded to all Black primes similarly. This effect could be explained by outgroup homogeneity (Linville, Fischer, & Salovey, 1989), the tendency to see outgroup members as more interchangeable while perceiving ingroup members as diverse. However, we when we explicated the prime gender × quadratic prototypicality × word type in Study 2 by prime gender, we also observed that the quadratic prototypicality × word type effect was significant for male prime trials. Follow up studies are needed to determine if this is a replicable pattern, and if it is, why it might be the case that prototypicality effects are typically only observed among some of the stimuli.

Features impact important, real-world judgments across multiple facets of life. As we describe at the outset of the paper, researchers have established that the one's facial physiognomy can affect how individuals are treated with respect to hugely consequential legal decisions (Blair, Judd, & Chapleau, 2004; Eberhardt et al., 2006; Ma & Correll, 2011). Others have illustrated that individual differences in stereotypical face features correlate with earnings, organizational prestige, and professional rank (Livingston & Pearce, 2009). Even social contact and rejection may relate to racial prototypicality. Researchers report that Blacks who look more prototypically Black have fewer non-Black friends and less interaction with outgroup individuals, and that more prototypic Blacks are subject to greater social rejection (Hebl, Williams, Sundermann, Kell, & Davies, 2012). Given the profound influence that features may have on individuals, gaining more clarity about the precise nature of how features impact judgment remains as important as ever.

References

Anderson, C., & Cromwell, R. L. (1977). "Black is beautiful" and the color preferences of Afro-American Youth. *The Journal of Negro Education, 46*, 76–88.

Bales, W. D. (1997). Sentencing guidelines: 1996–1997 annual report, Part II. *Impact, section 2. Sentencing neutrality*. (Retrieved July 2002 from the Florida Department of Corrections Web site) http://www.dc.state.fl.us/pub/sg_annual/9697/impact2.html.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211–227.

Blair, I. (2006). The efficient use of race and Afrocentric features in inverted faces. *Social Cognition, 24*, 563–579.

Blair, I., Chapleau, K., & Judd, C. (2005). The use of Afrocentric features as cues for judgment in the presence of diagnostic information. *European Journal of Social Psychology, 35*, 59–68.

Blair, I., & Judd, C. M. (2010). Afrocentric facial features and stereotyping. In R. B. AdamsJr., N. Ambady, K. Nakayama, & S. Shimojo (Eds.). *The science of social vision*. New York: Oxford University Press.

Blair, I., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*, 674–679.

Blair, I., Judd, C. M., & Fallman, J. L. (2004). The automaticity of race and Afrocentric facial features in social judgments. *Journal of Personality and Social Psychology, 87*, 763–778.

Blair, I., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: Judging by features and categories. *Journal of Personality and Social Psychology, 83*, 5–25.

Bodenhausen, G. V., & Wyer, R. S. (1987). Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.). *Social information processing and survey methodology* (pp. 6–41). New York: Springer Verlag.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin, 76*, 15–38.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin, 25*, 413–423.

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

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

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton

trilogy revisited. *Personality and Social Psychology Bulletin, 21*, 1139–1150.
Diekman, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin, 26*, 1171–1188.
Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology, 22*, 22–37.
Dunham, Y., Dotsch, R., Clark, A. R., & Stepanova, E. V. (2016). The development of white-Asian categorization: Contributions from skin color and other physiognomic cues. *PLoS One, 11*(6).
Dunham, Y., Stepanova, E. V., Dotsch, R., & Todorov, A. (2014). The development of race-based perceptual categorization: Skin color dominates early category judgments. *Developmental Science, 18*(3), 469–483.
Eagly, A. H., & Sczesny, S. (2009). *Stereotypes about women, men, and leaders: Have times changed?* American Psychological Association.
Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science, 17*, 383–386.
Faul, F., Erdfelder, E., Lang, A., & 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.
Hagiwara, N., Kashy, D. A., & Cesario, J. (2012). The independent effects of skin tone and facial features on Whites' affective reactions to blacks. *Journal of Experimental Social Psychology, 48*, 892–898.
Hamilton, D. L. (1976). Cognitive biases in the perception of social groups. In J. S. Carroll, & J. W. Payne (Eds.). *Cognition and social behavior*. Hillsdale, N.J: Lawrence Erlbaum Associates.
Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392–407.
Hebl, M. R., Williams, M. J., Sundermann, J. M., Kell, H. J., & Davies, P. G. (2012). Selectively friending: Racial stereotypicality and social rejection. *Journal of Experimental Social Psychology, 48*, 1329–1335.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135.
Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman, & J. A. Bargh (Eds.). *Unintended thought* (pp. 75–123). New York: Guilford Press.
Kahn, K. B., Unzueta, M. M., Davies, P. G., Alston, A. T., & Lee, J. K. (2015). Will you value me and do I value you? The effect of phenotypic racial stereotypicality on organizational evaluations. *Journal of Experimental Social Psychology, 59*, 130–138.
Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of ingroup and outgroup members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57*, 165–188. http://dx.doi.org/10.1037/0022-3514.57.2.165.
Livingston, R. W., & Brewer, M. B. (2002). What are we really priming?: Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.
Livingston, R. W., & Pearce, N. A. (2009). The teddy-bear effect. *Psychological Science, 20*, 1229–1236.
Ma, D. S., & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47*, 391–396.
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.
Ma, D. S., Correll, J., & Wittenbrink, B. (2016). Context dependency at recall: Decoupling context and targets at encoding. *Social Cognition, 34*, 119–132.
Maddox, K. B. (2004). Perspectives on racial phenotypicality bias. *Personality and Social Psychology Review,* (4), 383–401.
Maddox, K. B., & Gray, S. A. (2002). Cognitive representations of black Americans: Re-exploring the role of skin tone. *Personality and Social Psychology Bulletin, 28*, 250–259.
Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology, 106*, 226–254.
Niemann, Y. F., Jennings, L., Rozelle, R. M., Baxter, J. C., & Sullivan, E. (1994). Use of free responses and cluster analysis to determine stereotypes of eight groups. *Personality and Social Psychology Bulletin, 20*, 379–390.
Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*, 269–281.
Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12*, 1–20.
Ronquillo, J., Denson, T. F., Lickel, B., Lu, Z., Nandy, A., & Maddox, K. B. (2007). The effects of skin tone on race-related amygdala activity: An fMRI investigation. *Social Cognitive and Affective Neuroscience, 2*, 39–44.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 16*, 371–416.
Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.
Sassenberg, K., & Wieber, F. (2005). Don't ignore the other half: The impact of ingroup identification on implicit measures of prejudice. *European Journal of Social Psychology, 35*, 621–632.
Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: Assimilation and contrast effects in social judgment. In L. L. Martin, & A. Tesser (Eds.). *The construction of social judgment* (pp. 217–245). Hillsdale, NJ: Erlbaum.
Shapiro, J. R., Mistler, S. A., & Neuberg, S. L. (2010). Threatened selves and differential prejudice expression by white and black perceivers. *Journal of Experimental Social Psychology, 46*, 469–473.
Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*, 214–241.
Spiers, H. J., Love, B. C., Le Pelley, M. E., Gibb, C. E., & Murphy, R. A. (2017). Anterior temporal lobe tracks the formation of prejudice. *Journal of Cognitive Neuroscience, 29*, 530–544.
Stepanova, E. V., & Strube, M. J. (2012). The role of skin color and facial physiognomy in racial categorization: Moderation by implicit racial attitudes. *Journal of Experimental Social Psychology, 48*(4), 867–878.
Uhlmann, E., Dasgupta, N., Elgueta, A., Greenwald, A. G., & Swanson, J. (2002). Subgroup prejudice base on skin tone among Hispanics in the United States and Latin America. *Social Cognition, 20*, 198–225.
Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.
Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology, 37*, 244–252.
Wittenbrink, B., & Schwarz, B. (Eds.). (2007). *Implicit measures of attitudes*. New York: Guilford Press.