Taking Priming to Task
Variations in Stereotype Priming Effects Across Participant Task

Katherine R. G. White,1 Rose H. Danek,2 David R. Herring,3 Jennifer H. Taylor,4 and Stephen L. Crites5

1Department of Psychology, Kennesaw State University, GA, USA
2Department of Psychology, Lyon College, Batesville, AR, USA
3Department of Psychology, Pennsylvania State University at Erie, PA, USA
4Department of Psychology, Lock Haven University, PA, USA
5Department of Psychology, University of Texas at El Paso, TX, USA

Abstract: The current research examined potential moderators of gender and racial stereotype priming in sequential priming paradigms. Results from five experiments suggest that stereotype priming effects are more consistent in tasks that elicit both semantic priming and response competition (i.e., response priming paradigms) rather than tasks that evoke semantic priming alone (i.e., semantic priming paradigms). Recommendations for future stereotype priming research and the implication of these results for the proper interpretation of stereotype priming effects are discussed.

Keywords: stereotypes, priming, participant task, gender, race

Priming occurs when a stimulus or context influences subsequent cognitions and behavior. Priming has expanded from initial research on cognition and memory (e.g., Schacter, Dobbins, & Schnyer, 2004) to topics such as stereotyping and attitudes (see Wittenbrink, 2007). A popular paradigm for studying stereotype priming is the sequential priming paradigm. In a typical sequential priming paradigm (see McNamara, 2005; Neely, 1977), participants are presented trials containing two stimuli – a prime and target – and prime-target congruency is varied across trials. For example, a stereotypically congruent trial might be the prime “WOMEN” followed by the target “NURSE.” A stereotypically incongruent trial might be the prime “MEN” followed by the target “NURSE.” A stereotype priming effect occurs when participants’ responses to targets are faster and more accurate on stereotypically congruent than incongruent trials. While researchers have identified numerous variables that moderate this effect (for review, see Blair, 2002), some inconsistencies remain unexplained. This paper outlines a set of studies (conducted over an eight-year period) which attempted to identify variables that might explain inconsistent findings in sequential stereotype priming.

Our investigation began with research that explored whether the N400 event-related potential (ERP) component could measure stereotype priming (White, Crites, Taylor, & Corral, 2009). In this study, gender categories “MEN” and “WOMEN” served as primes and participants were asked to indicate whether each target word (e.g., PURSE, MUSCLES) matched or did not match the prime according to common gender stereotypes. As expected, response times were slower and N400 amplitude was larger when targets did NOT match the preceding prime. In this initial study, we intentionally used an explicit decision task to enhance the N400 effect (Chwilla & Kolk, 2005). However, people do not typically make explicit stereotype judgments when they encounter other individuals, which makes a matching task less theoretically interesting than tasks that more directly assess “automatic” stereotype activation. We therefore performed a follow-up study in which participants indicated whether each target was a word or nonword (i.e., lexical decision task or LDT). Evidence of priming with this type of task would suggest that stereotype priming occurs even when people are engaged in a task that is irrelevant to the stereotype dimension under investigation. However, stereotype priming failed to manifest for both response times and the N400 (White & Crites, 2009). We then performed an additional small pilot study (just focused on response times), and this study again found no evidence of stereotype priming. Neither of these LDT studies are reported in detail here.
because they were potentially underpowered, but their results prompted us to a focus on participant task as a potential moderator.

Participants can perform many different judgment tasks in a sequential priming paradigm. Some tasks used in previous stereotype priming research include the LDT, Stroop-pronunciation task, semantic classification, and stereotype classification task (SCT). The LDT requires participants to categorize each target as a word (e.g., NURSE) or nonword (e.g., NIRSE). In the pronunciation task, participants must simply read the target word out loud. Semantic classification tasks require participants to categorize targets into one of two semantic categories (e.g., person/place, noun/pronoun) that are presumably irrelevant to the stereotype dimension being examined. The SCT requires participants to categorize target stimuli along a stereotype-relevant dimension (e.g., categorize targets as male or female when gender stereotypes are under investigation, or black-white for race stereotypes). We broadly divided these various tasks into two categories—those that require a stereotype-relevant judgment about the target under investigation (SCT) and those that require a stereotype-irrelevant judgment about the target (LDT, pronunciation, semantic classification).

An interesting pattern seems to emerge when examining the judgment tasks used in previous stereotype priming studies (see Table 1). Significant stereotype priming effects are consistently observed when a stereotype-relevant classification task is used, but are less consistent when a stereotype-irrelevant task is used. Notably, no study has closely examined the effect of participant task on the stereotype priming effect. One exception is the article by Banaji and Hardin (1996), which observed stronger gender stereotype priming when participants categorized targets as male/female (i.e., an SCT) versus pronoun/not pronoun (i.e., a semantic classification). Banaji and Hardin (1996) state that this finding requires further exploration in future research, but to our knowledge, this has not been done. In this article, we report results from five experiments which systematically examine the potential impact of participant task on the stereotype priming effect. These studies took place over 8 years, during which time theoretical explanations that closely align with our pattern of findings were also being articulated (Wentura & Degner, 2010; Wentura & Rothermund, 2014). We explore these theoretical explanations in the discussion and here present the five experiments we conducted in an effort to better understand the inconsistencies observed in our own and others’ research.

### Study 1

This study examines the impact of participant task on gender stereotype priming. Participants were randomly assigned to either a stereotype classification task (SCT condition) or a lexical decision task (LDT condition). Participants in the SCT condition categorized targets as more associated with men or women. Participants in the LDT condition categorized targets as words or nonwords. We also randomly assigned participants to a third task condition—the pre-primed LDT condition. In this condition, participants completed an LDT after engaging in a short task to make gender more salient/accessible (see Wittenbrink et al., 1997 for similar procedure). We included this pre-primed LDT condition to explore the possibility that stereotype priming can be boosted by making stereotypes temporarily more accessible (Higgins & King, 1981). We predicted that gender stereotype priming would be significant in both the SCT and pre-primed LDT conditions, but stronger in the SCT condition.

### Method

#### Participants

The final sample included 171 participants: 52 in gender categorization, 59 in classic LDT, and 60 in pre-primed LDT. A total of 199 (146 female) undergraduates participated for partial course credit; however, data from 28 participants were not available for analyses or excluded from the primary analyses. In the entire sample, 169 participants self-identified as Hispanic/Latino. Ages ranged from 18 to 46 years (M = 19.0).

#### Stimuli

Trials consisted of a prime followed by a target. Trials were either a stereotype congruent pair (e.g., STEPHANIE: GOSSIPY), a stereotype incongruent pair (e.g., JOSEPH: GOSSIPY), or a pair with a nonword target (e.g., MARIO: EEB4). Nonword targets only appeared in the LDT and pre-primed LDT conditions.

---

1. As discussed above, this research began in 2008. To help place these studies in context, we report here when each was initiated and completed. Study 1 was initiated and completed in the fall of 2009. Study 2 took place in the spring of 2011 and Study 3 in the spring of 2013. Finally, Studies 4 and 5 were initiated in the spring of 2015 and completed in the spring of 2016.

2. We used an LDT as an exemplar of a stereotype-irrelevant task because it is more commonly used in stereotype priming research and to maintain consistency with our previous LDT studies which failed to produce significant stereotype priming.

3. See Appendix B for discussion of how (1) data were processed prior to analyses, (2) this led to data from participants being excluded, and (3) additional analyses that verify those reported in the text.
Primes
Primes were 54 male names, 54 female names, 54 pictures of males, and 54 pictures of females. An additional eight names and eight pictures were used in practice trials.

Targets
Targets were either words associated with gender or nonwords. The gender words consisted of 16 female and 16 male stereotypical traits or nontraits. Stereotype words were matched for valence and length and normed in the local population (see Appendix A for a complete list of stimuli and norming results). Twenty-four additional words (12 male and 12 female) associated with gender were used in practice trials. Nonwords were 56 pronounceable nonwords from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002).
Procedure

Participants were run in small groups, with each group assigned to one of the three conditions. After informed consent, participants completed a short demographics questionnaire and were told that they would see a series of stimulus pairs on the computer. They were asked to attend to the first stimulus (prime) and respond to the second (target) as quickly and accurately as possible, using designated keys on a computer keyboard.

In the SCT condition, targets were always words and participants indicated whether each was associated with men or women—half were stereotypically congruent with the prime and half were stereotypically incongruent. In the LDT and pre-primed LDT conditions, half of the targets were words and half were nonwords. Participants indicated whether each was a word or nonword. For word targets, half were stereotypically congruent with the prime and half were stereotypically incongruent.

The pre-priming session for the pre-primed LDT condition consisted of 24 names (12 female and 12 male) and 24 pictures (12 female and 12 male) that participants categorized as male or female. The order of names/pictures was counterbalanced across participants. The 48 stimuli in this pre-priming session were not used in the experimental session. Following the pre-priming session, participants initiated the LDT.

For all conditions, trials consisted of a 200 ms focus “+”, a 150 ms prime, a 100 ms blank screen (stimulus-onset asynchrony, SOA = 250 ms), and a target that lasted until the participant responded but no more than 1,050 ms. There was a 1,500 ms intertrial interval (ITI) before the onset of the focus for the next trial. All stimuli were presented on a black background. Verbal stimuli were white capital letters.

Trials were organized into seven blocks. The first consisted of 48 practice trials, after which participants could ask the experimenter if they had questions. The six experimental blocks each began with eight practice trials (i.e., warm-up trials; Wentura & Degner, 2010) followed by 96 experimental trials. Three experimental blocks used picture primes and three used name primes. The picture and name blocks alternated and were counterbalanced across participants. Following the priming procedure, all participants rated the masculinity/femininity (1 = very masculine, 7 = very feminine) of all targets. Male target stimuli were perceived by participants as significantly more masculine ($M_M = -1.97$) than female target stimuli ($M_F = 1.91$), $t = -20.92, p < .001$.

Results and Discussion

Response times from correct responses were analyzed with a 3 Condition: gender categorization, LDT, or pre-primed LDT) × 2 Congruency: stereotype congruent vs. incongruent mixed factorial analysis of variance (ANOVA) with Condition between-subjects. There were main effects for both Congruency, $F(1, 168) = 66.75, p < .001$, $\eta^2 = .284$, 90% CI [.192, .369], and Condition, $F(2, 168) = 10.18, p < .001$, $\eta^2 = .108$, 90% CI [.040, .179]. These were qualified by a significant Condition by Congruency interaction, $F(2, 168) = 37.20, p < .001$, $\eta^2 = .307$, 90% CI [.209, .387]. As expected, stereotype priming was significant in the SCT condition as people responded faster to congruent ($M = 603$ ms; $SD = 55.3$) than incongruent ($M = 623$ ms; $SD = 54.0$) targets, $F(1, 51) = 76.99, p < .001$, $\eta^2 = .602$, 90% CI [.449, .692]. Stereotype priming was nonsignificant in the LDT ($M_s = 563$ and 565 ms; $SD_s = 60.9$ and 62.5 for congruent and incongruent, respectively), $F(1, 58) = 2.33, p = .132$, $\eta^2 = .039$, 90% CI [.00, .144], or pre-primed LDT conditions ($M_s = 570$ and 571 ms; $SD_s = 68.6$ and 68.4 for congruent and incongruent, respectively), $F(1, 59) = 1.45, p = .233$, $\eta^2 = .024$, 90% CI [.00, .118].

Results provide initial evidence that participant task is a significant moderator of the stereotype priming effect and that increasing the salience of the stereotype dimension (via pre-priming) does not lead to greater stereotype priming in stereotype-irrelevant tasks. Specifically, gender stereotype priming was significant when participants categorized targets as male/female (i.e., SCT) but not when participants categorized targets as word/nonword (i.e., LDT) in either the traditional or pre-primed LDTs. We speculated that past inconsistencies in stereotype-irrelevant tasks may have been due to extraneous factors that lead to stereotypes being more salient in some studies (which found priming) and less salient in others (which did not find priming). Thus, we attempted to increase stereotype salience in the pre-primed LDT to test this idea.

---

4 The eight practice trials preceding each experimental block were not separated from the experimental trials and thus appeared to be experimental trials from the participants’ perspective. They were used to get participants into the flow of the experiment after a break. There were eight trials in order to present all four trial types (e.g., male-male; male-female; female-female; female-male) and an equivalent number of non-word trials for the LDT tasks. The SCT task used two presentations of the four trial types to maintain an equivalent number of practice trials.

5 Picture versus name primes were used to explore whether this impacted magnitude of priming. Initial analyses suggested that it did not; so it will not be discussed further.

6 Because the SCT does not have non-word trials, it has twice the number of available trials for analyses as an LDT. Thus, we examined the SCT using all trials and half of the trials (interpolated through the entire range of the task to be comparable with the word trials found in an LDT). There was no difference in the pattern of significance in this Study or Studies 3 and 5, which also compared SCT to LDT. Thus, in all three studies, we report analyses that include all SCT trials.
(i.e., by comparing priming in traditional vs. pre-primed LDTs). It is possible that the pre-priming procedure used in Study 1 was not strong enough to boost the salience of gender across the entire LDT. We therefore explored a different method of increasing stereotype salience in Study 2.

### Study 2

An advantage of stereotype classification tasks is that they focus participants on the stereotype dimension throughout the entire task. One limitation of Study 1’s pre-priming procedure is that any increase in the salience of gender may have quickly dissipated. Recent research suggests that evaluative priming may only be observed in conditions where attention is directed to the evaluative dimension, referred to as the attention allocation perspective (Spruyt, De Houwer, Everaert, & Hermans, 2012; Spruyt, De Houwer, & Hermans, 2009). This type of attention moderation may also occur for stereotype priming.

In Study 2, we adapted procedures previously used to test the attention allocation perspective to design a modified LDT that would make gender salient throughout the LDT (see Spruyt, De Houwer, Hermans, & Eelen, 2007, Study 2 for similar procedure). Participants were randomly assigned to either count the number of male and female primes throughout the LDT (gender tally condition) or the number of Hispanic and white primes throughout the LDT (ethnicity tally condition). We predicted that the increased salience of gender in the gender tally condition would be sufficient to create gender stereotype priming in the LDT. This would provide one potential explanation for prior inconsistencies in stereotype priming results while simultaneously adding support to the attention allocation perspective.

### Method

#### Participants

The final sample included data from 121 participants: 53 in the ethnicity tally and 68 in the gender tally. A total of 149 (90 female) undergraduates participated for partial course credit; however, data from 28 participants were not available for analyses or excluded from the primary analyses (see Appendix B). One hundred thirty-two participants self-identified as Hispanic/Latino. Ages ranged from 18 to 45 years ($M = 21.0$).

#### Stimuli

Trials consisted of a prime followed by a target and could form a stereotype congruent pair, stereotype incongruent pair, or a pair with a nonword target. The prime names (54 male and 54 female), target words (16 male and 16 female), target nonwords, and practice stimuli were the same as those used in Study 1. Unlike Study 1, picture primes were not used.

#### Procedure

Participants were run in small groups, with each group assigned to one of the two tally conditions. Following a demographic questionnaire, participants were told that the experiment consisted of a series of trials in which they would see a name followed quickly by a word or nonword. The experimenter explained that they would have slightly different tasks across three phases of the experiment.

##### Phase 1

Phase 1 was a practice phase to introduce participants to the LDT. All participants categorized each target as a word or nonword as quickly and accurately as possible, using two keys on a computer keyboard. The nature and timing of this LDT was identical to that in Study 1 except that there were two blocks of trials that each contained 16 trials.

##### Phase 2

Phase 2 was also a practice phase. Participants’ primary task was to categorize each target as a word or nonword as quickly and accurately as possible. Their secondary task was to mentally tally either the number of male and female names that appeared (gender tally) or the number of Hispanic and white names that appeared (ethnicity tally) and report this number at the end of each block. The experimenter stressed that the lexical decision task was the primary task and the mental tally was secondary. The timing of each trial was identical to those in Phase 1 except that after a response to a target, a screen reminded participants to update their mental tally of male and female (or Hispanic and white) names. Phase 2 consisted of six blocks of 16 trials, and participants entered their tally at the end of each block.

##### Phase 3

Phase 3 contained the critical experimental trials and was similar to Phase 2 except there was no reminder to update the mental tally after each trial. A 3,000 ms ITI was used between trials. Phase 3 consisted of 18 experimental blocks of 16 trials each. At the end of each block, participants reported their tally from the previous block. Experimental blocks were created with varying numbers of male/female (or Hispanic/white) primes. Of the 18 blocks, 6 contained 8 male and 8 female names, 2 contained 9 male and 7 female names, 2 contained 6 male and 10 female names, and so forth. Across all 18 blocks, the number of trials with male primes was the same as the number of trials with female primes (144 trials each).
Results and Discussion

Response times from Phase 3 correct responses were analyzed with a 2 (Tally: gender vs. ethnicity) × 2 (Congruency: stereotype congruent vs. incongruent) mixed factorial ANOVA with Condition between-subjects. Neither the main effect of Congruency (Ms = 687 and 690 ms; SDs = 83.9 and 83.6 for congruent and incongruent, respectively), F(1, 119) = 2.02, p = .158, η² = .017, 90% CI [.00, .072], nor the interaction between Tally and Congruency, F(1, 119) = 0.01, p = .933, η² < .001, 90% CI [.00, .004], was significant. Because we predicted priming in the gender tally condition, we performed simple effect analyses to examine congruency in each condition. The congruency effect was nonsignificant in both the ethnicity tally condition, F(1, 52) = .78, p = .38, η² = .015, 90% CI [.00, .106], and the gender tally condition, F (1, 67) = 1.31, p = .256, η² = .019, 90% CI [.00, .102] (see Appendix B for supplementary analyses).

Gender stereotype priming was absent in both the gender and ethnicity tally conditions. Thus, increased attention to the stereotype dimension of interest throughout the LDT was not sufficient to elicit gender stereotype priming. When combined with Study 1’s results, this suggests that gender stereotype priming effects are difficult to detect when participants perform an LDT. Study 3 examined whether this is true for another commonly researched stereotype dimension – race.

Study 3

Results from Studies 1 and 2 suggest that stereotype priming effects are difficult to capture when participants perform an LDT and also suggest that attention allocation may not adequately explain inconsistencies in previous stereotype priming research. However, both studies examined gender stereotypes. Study 3 examined whether these results generalize to race stereotype priming. As seen in Table 1, half (50%) of the published studies on gender stereotypes that used a stereotype-irrelevant task (i.e., LDT, pronunciation, or semantic classification) report nonsignificant results. This is greater than the proportion of null results reported for race stereotypes (≈ 21%). This suggests that race, relative to gender, stereotype priming may be more easily observed in stereotype-irrelevant tasks.

Study 3 examined race stereotypes and randomly assigned participants to complete either a lexical decision (LDT)7 or stereotype classification (SCT) condition. Given the previous pattern of results (see Table 1), we expected to observe significant race stereotype priming in the SCT condition as well as significant, albeit weaker, race stereotype priming in the LDT condition.

Method

Participants

The final sample included 119 participants: 76 in the LDT and 43 in the SCT. A total of 150 undergraduates participated; however, data from 31 participants were not available for analyses or excluded from the primary analyses (see Appendix B).8 One hundred thirty-two participants self-identified as Hispanic/Latino. Ages ranged from 18 to 45 years (M = 21.0).

Stimuli

Trials consisted of a prime stimulus followed by a target stimulus and could form a stereotype congruent pair (e.g., HISPANIC PICTURE: SOCCER), stereotype incongruent pair (e.g., BLACK PICTURE: SOCCER), or a pair with a nonword target (e.g., HISPANIC PICTURE: SPOMPH).

Primes

Prime stimuli consisted of 80 pictures – 40 of Hispanics (20 male and 20 female) and 40 of Blacks (20 male and 20 female). An additional 20 pictures (10 Hispanic and 10 Black) were used for practice trials.

Targets

Targets were 20 words stereotypically associated with blacks (e.g., basketball, trendy), 20 words stereotypically associated with Hispanics (e.g., agriculture, immigration), and 80 pronounceable nonwords obtained from the ARC Nonword Database generator (Rastle et al., 2002; see Appendix A for complete list). An additional 12 words of each type were used in practice trials.

Procedure

The procedure was identical to Study 1, which also had both LDT and SCT conditions, except for the following

---

7 We continued to use an LDT for consistency with our previous studies.
8 In the initial computer program, all target words were in "CAPS" but non-words were in "lowercase." This problem, which only affected the LDT condition, was discovered after data from 17 participants were collected in the LDT condition. We collected additional participants in the LDT condition because we anticipated having to remove these 17 participants. That is, this CAP versus lowercase confound meant that once participants realized this difference, they would not have to read the target stimuli and could respond word/non-word based solely on the form (i.e., CAP = "word" vs. lowercase = "non-word"). The pattern of results and significance were identical when including or not including these participants. So, we report analyses that include these 17 participants.
differences. First, there was no pre-primed LDT condition. Second, only picture primes were used. Third, in the SCT, participants indicated whether each target was more associated with blacks (e.g., BASKETBALL) or Hispanics (e.g., SOCCER), rather than men or women. Fourth, each trial consisted of a 650 ms focus “+”, a 50 ms blank screen, a 200 ms prime, another 50 ms blank screen, and a target that lasted until the participant responded but no more than 1,050 ms. Fifth, trials were organized into eight blocks. The first consisted of 24 practice trials. The following seven experimental blocks each began with four practice trials followed by 80 experimental trials.9

Results and Discussion

Response times from correct trials were analyzed with a 2 (Condition: LDT vs. SCT) × 2 (Congruency: stereotype congruent vs. incongruent) mixed factorial ANOVA with Condition between-subjects (see Appendix B). This analysis revealed the expected Congruency main effect, $F(1, 109) = 49.67$, $p < .001$, $\eta^2 = .313$, 90% CI [.197, .414], and the Congruency by Condition interaction, $F(1, 109) = 30.70$, $p < .001$, $\eta^2 = .219$, 90% CI [.114, .324]. Follow-up analyses revealed a significant congruency effect in the SCT condition, $F(1, 42) = 27.92$, $p < .001$, $\eta^2 = .399$, 90% CI [.204, .538]. Participants responded faster to congruent ($M = 644$ ms; $SD = 55.8$) than incongruent ($M = 668$ ms; $SD = 52.9$) targets. Participants in the LDT condition also responded more quickly to congruent ($M = 677$ ms; $SD = 58.0$) than incongruent ($M = 682$ ms; $SD = 58.0$) targets, $F(1, 67) = 10.20$, $p = .002$, $\eta^2 = .132$, 90% CI [.031, .257]. The difference between incongruent and congruent trials in the SCT ($M = 23.6$ ms; $SD = 29.3$) is significantly greater than that in the LDT ($M = 5.1$ ms; $SD = 13.1$), $F(1, 109) = 20.70$, $p < .001$, $\eta^2 = .160$, 90% CI [.067, .261].10

Matching our predictions, race stereotype priming was observed when participants completed a lexical decision task. Additionally, priming in the LDT condition was weaker than in the SCT condition. This strengthens the assertion that stereotype priming effects are robust when participants perform an SCT. When combined with results from Studies 1 and 2, the results suggest that stereotype priming effects are less consistent when participants perform an LDT. They also raise the possibility that stereotype dimension moderates the stereotype priming effect. Studies 4 and 5 simultaneously investigated this possibility.

Study 4

Studies 1 and 3 provide evidence of robust stereotype priming when participants perform an SCT. Evidence for stereotype priming when participants perform an LDT, however, was observed for racial stereotypes but not for gender stereotypes. Study 4 directly tested stereotype dimension as a moderator of stereotype priming when participants perform an LDT. All participants performed an LDT but were randomly assigned to either a race stereotypes or gender stereotypes condition. We hypothesized that significant stereotype priming effects would emerge for race but not for gender.

Method

Participants

The final sample included 165 participants: 87 in the race stereotypes condition and 78 in the gender stereotypes condition. A total of 178 undergraduates participated; however, data from 13 participants were not available for analyses or excluded from the primary analyses (see Appendix B). One hundred fifty-six self-identified as Hispanic/Latino. Ages ranged from 18 to 53 years ($M = 20.8$).

Stimuli

Trials consisted of a prime followed by a target and were either a stereotype congruent pair, a stereotype incongruent pair, or a pair with a nonword target.

Primes

The pictures used for experimental (20 black females, 20 Hispanic females, 20 black males, and 20 Hispanic males) and practice primes were the same as those in Study 3.

Targets

Targets were a set of 80 words and 80 pronounceable nonwords from the ARC Nonword Database (Rastle et al., 2002). Words consisted of 20 words from each of the four stereotype categories (female, male, Black, and Hispanic). In the race condition, the target words were 20 black and 20 Hispanic stereotype words. In the gender condition, the target words were 20 female and 20 male stereotype words. Words were matched for valence and length and normed in the local population (see Appendix A).

---

9 As in Study 1, we examined the SCT using all trials and half of the trials and found no difference in the pattern of significance in these two analyses. We report the analysis that includes all SCT trials.

10 When we analyzed data from 137 participants (of the 150 total) who had data in all experimental conditions, the global analyses revealed the same Congruency main effect and Condition by Congruency interaction. The congruency effect for the SCT remained significant, but the effect for the LDT was not, $F(1, 74) = 3.08$, $p = .083$, $\eta^2 = .040$. 

© 2018 Hogrefe Publishing
Procedure
The procedure was identical to the LDT task used in Study 3 except for the following differences. First, participants were assigned to either the gender or race condition. In the gender condition, half the word trials were stereotype congruent (e.g., male picture: male stereotype word) and half were stereotype incongruent (e.g., male picture: female stereotype word). In the race condition, half the word trials were stereotype congruent (e.g., black picture: black stereotype word) and half were stereotype incongruent (e.g., black picture: Hispanic stereotype word). Second, the ITI was 1,200 ms not 1,500 ms. Third, trials were organized into 7 blocks that included a block of 24 practice trials and 6 experimental blocks each began with 4 practice trials followed by 80 experimental trials.

Results and Discussion
Response times from correct responses were analyzed with a 2 (Condition: Race vs. Gender) × 2 (Congruency: stereotype congruent vs. incongruent) mixed factorial ANOVA with Condition between-subjects. This revealed nonsignificant effects for both the Congruency main effect (Ms = 574 and 575 ms; SDs = 56.2 and 56.1 for congruent and incongruent, respectively), F(1, 163) = 0.24, p = .626, η² = .002, 90% CI [.000, .026], and the Condition by Congruency interaction, F(1, 163) = 0.51, p = .477, η² = .003, 90% CI [.000, .032]. Because we predicted a significant finding for race and not gender, we conducted follow-up analyses. These analyses revealed no significant priming for either race (Ms = 582 and 583 ms; SDs = 51.3 and 51.5 for congruent and incongruent, respectively), F(1, 86) = 0.62, p = .433, η² = .007, 90% CI [.000, .063], or gender (Ms = 566 and 566 ms; SDs = 60.5 and 60.0 for congruent and incongruent, respectively), F(1, 77) = 0.32, p = .858, η² = .004, 90% CI [.000, .057].

Contrary to predictions and results from Study 3, stereotype priming failed to reach significance for both race and gender stereotypes. This once again underscores the inconsistent nature of stereotype priming effects when an LDT is used. While Study 4 was being conducted at a university in the Southwest that has a predominantly Hispanic student demographic, we were simultaneously attempting to replicate the findings of Study 3 at another university in the Southeast with a substantial black population.

Study 5
In this study, we manipulated both participant task (SCT vs. LDT) and stereotype dimension (race vs. gender). We used both an SCT and LDT in this study because it was conducted at a new university with different participant demographics. Participants were randomly assigned to one of four conditions—Race-LDT, Race-SCT, Gender-LDT, or Gender-SCT. We predicted that stereotype priming would be significant in both SCT conditions, significant but weaker in the Race-LDT condition, and absent in the Gender-LDT condition.

Method
Participants
The final sample included 217 participants (76.1% female) from a midsized Southeastern university (M = 21.84 years): 60 in the Race-LDT condition, 46 in the Race-SCT condition, 54 in the Gender-LDT condition, and 57 in the Gender-SCT condition. A total of 238 undergraduates participated; however, data from 21 participants were not available for analyses or excluded from the primary analyses (see Appendix B). Forty-one percent self-identified as White/Caucasian, 43.2% as Black/African American, 7.7% as Hispanic/Latino, and 9.9% as Pacific Islander, American Indian, Asian, multiracial, or said they preferred not to answer.

Stimuli
Trials consisted of a prime stimulus followed by a target stimulus. For participants assigned to one of the two Race conditions, trials were race stereotype-congruent, race stereotype-incongruent, or had a nonword target. Participants assigned to one of the two Gender conditions saw trials that were gender stereotype-congruent, gender stereotype-incongruent, or had a nonword target. Nonword targets only appeared in the LDT conditions.

Primes
Prime stimuli consisted of 120 pictures—30 black females, 30 black males, 30 white females, and 30 white males. Pictures were from the Park Aging Mind Laboratory (Minear & Park, 2004) or were used courtesy of Michael J. Tarr (Righi, Peissig, & Tarr, 2012). Pictures were headshots on a neutral or white background. Pictures were previously rated as neutral or near neutral in attractiveness. An additional five pictures of each type were used for practice trials only.

Targets
Targets were words associated with race, gender, or nonwords. Race words consisted of 30 words stereotypically associated with blacks and 30 words stereotypically associated with whites (see Appendix A for complete list). Gender words consisted of 30 words stereotypically associated with men and 30 words stereotypically associated with women. An additional eight words of each type were used for practice trials only.
Both race and gender stimuli were matched for valence and normed in the local population (see Appendix A). Nonwords were 60 pronounceable nonwords created by changing two letters in the word stimuli (e.g., BASPETHALL), rather than words from the ARC Nonword Database. Ten additional nonwords were created and used in practice trials.

Procedure
The procedure was identical to Study 3, which also had both LDT and SCT procedures, except for the following differences. First, participants were randomly assigned to one of the four conditions (Race-LDT, Race-Categorization, Gender-LDT, Gender-Categorization). Participants in the Race-SCT were instructed to indicate whether the target was more associated with blacks or whites. Participants in the Gender-SCT were instructed to indicate whether the target was more associated with men or women. Second, the target lasted until the participant responded but no more than 1,500 ms and the ITI was 1,200 ms.

Results and Discussion
The omnibus 2 (Task: LDT, categorization) × 2 (Stereotype Dimension: race, gender) × 2 (Congruency: stereotype congruent vs. incongruent) ANOVA revealed significant effects for all main effects and interactions, including a Task × Stereotype Dimension × Congruency interaction, F(1, 213) = 7.87, p = .005, η² = .036. This significant three-way interaction suggests that the critical within-subject congruency effect varies across the four unique between-subject conditions. We probed this by testing the congruency effect in each condition (see Table 2).

Table 2. Mean, standard deviation, and p-values for congruent and incongruent trials in Study 5

<table>
<thead>
<tr>
<th>Condition</th>
<th>Congruent M (SD)</th>
<th>Incongruent M (SD)</th>
<th>Significance (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race – SCT</td>
<td>759 (83.5)</td>
<td>776 (88.2)</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Gender – SCT</td>
<td>689 (75.0)</td>
<td>690 (75.6)</td>
<td>ns</td>
</tr>
<tr>
<td>Race – LDT</td>
<td>691 (87.1)</td>
<td>694 (88.5)</td>
<td>ns</td>
</tr>
<tr>
<td>Gender – LDT</td>
<td>683 (79.8)</td>
<td>683 (79.7)</td>
<td>ns</td>
</tr>
</tbody>
</table>

As expected, participants responded faster to congruent than incongruent targets in the Race-SCT, F(1, 45) = 15.94, p < .001, η² = .262, 90% CI [.093, .414]. However, there was no evidence of stereotype priming in the Race-LDT, F(1, 59) = 1.56, p = .216, η² = .026, 90% CI [.00, .121], Gender-SCT, F(1, 56) = 0.16, p = .695, η² = .003, 90% CI [.00, .062], or the Gender-LDT, F(1, 53) = 0.002, p = .970, η² < .001, 90% CI [.00, .002]. Thus, the results reveal significant stereotype priming only for race in a stereotype categorization task.

General Discussion
Taken as a whole, results from the five experiments indicate that participant task is a significant moderator of the stereotype priming effect. When participants performed a stereotype classification task (SCT), significant stereotype priming was observed with relatively high consistency in three of four within-subject tests across both gender and race stereotypes (note that Study 5 has two “tests,” one for gender and another for race). By contrast, stereotype priming was largely absent when participants performed a lexical decision task (LDT), reaching significance only once. To further explore this issue, we completed a meta-analysis of the five studies reported here and included one pilot study referenced in the Introduction. Using a random effects model, we found that stereotype priming was significant overall (d = 0.20, SE = 0.065, p = .002, 95% CI [.07, .33]), but there was also significant heterogeneity among the studies, Q(10) = 24.14, p = .007. To test whether the heterogeneity was due to task, we ran a mixed effects model with task as a categorical moderator. The model accounted for 58.33% of the differences in effect sizes between studies, QM(I) = 4.73, p = .030. While stereotype priming was significant for both the stereotype categorization (k = 3, d = 0.42, SE = 0.12, p < .001, 95% CI [.19, .66]) and lexical decision (k = 8, d = 0.13, SE = 0.07, p = .049, 95% CI [.05, .26]) tasks, priming was significantly larger for the SCT than the LDT tasks, χ² = 4.73, p = .030. These results largely mirror those reported in a full-scale meta-analysis conducted by our laboratory (Kidder, White, Hinojos, Sandoval, & Crites, in press), with the exception that the stereotype priming effect for the LDT failed to reach significance in the full meta-analysis.

As discussed earlier, we began this line of research in 2009 as a follow-up to a study that demonstrated that the N400 component of the event-related potential (ERP) is associated with stereotype priming with a task conceptually similar to the SCT (White et al., 2009). As we were conducting these studies, other researchers (Wentura & Degner, 2010; Wentura & Rothermund, 2014) were articulating a theoretical framework that can help explain the pattern of results across our studies.

Wentura and Degner (2010) divide sequential priming paradigms into two categories – semantic and response priming paradigms. Differentiating between these paradigms is important because different cognitive mechanisms are thought to underlie priming effects observed in these different paradigms (Wentura & Rothermund, 2014). A critical variable that distinguishes these two types of paradigms is the judgment that people make about target stimuli. Sequential priming designs that use an LDT, pronunciation, or semantic classification task fall under the category of...
Social Psychology

targets. These paradigms are thought to be driven primarily by the priming paradigms because priming observed in semantic priming paradigms is traditionally attributed to spreading activation (Collins & Loftus, 1975) or the activation of overlapping mental representations (e.g., Masson, 1995; Rumelhart & McClelland, 1986) in memory. For example, seeing a female picture prime activates the concept of “female” in memory, which then facilitates responses to other items which share those semantic features. As “nurse” is stereotypically female, the pattern of activation for nurse is facilitated by the preceding concept of “female.” Sequential priming paradigms that use a task like the SCT fall under the category of response priming paradigms because the judgment about targets can also be applied to primes, which may create facilitation/competition for target responses. Priming effects observed in response priming paradigms are attributed to both memory activation and response activation that occurs when a prime activates a response that is either compatible or incompatible with the response required by the target. For example, people in a gender stereotype priming SCT have to respond “female” when they see the target word “nurse.” Seeing a female name as the prime activates the response “female” to that prime, which then makes it easier to respond “female” to the word “nurse.” Alternatively, seeing a male name as the prime activates the response “male,” which interferes with the necessary response to the target word “nurse.” It is important to note that response priming designs may involve BOTH semantic and response priming whereas semantic priming designs involve only semantic priming (although see Footnote 11). Thus, Wentura and Degner (2010) assert that priming effects in response priming paradigms may be more robust than those observed in semantic priming paradigms.

The present research results align nicely with this framework. On several occasions, Wentura and Degner (2010) suggest that response priming effects are relatively robust (although not entirely consistent) and that semantic priming effects tend to be more mercurial. With only one exception in the present set of studies (Study 5, gender), significant stereotype priming was consistently observed when an SCT (response priming paradigm) was used. Moreover, the meta-analysis of these results demonstrated that the stereotype priming effect observed with an SCT is significantly stronger than priming effects observed with an LDT. Priming effects in response priming paradigms likely reflect the combined influence of semantic and response priming processes (Wentura & Rothermund, 2014). The combined influence of these two processes ostensibly makes for a stronger, more robust stereotype priming effect in response priming paradigms. This can explain the more consistent (and stronger) pattern of significant stereotype priming effects when participants performed an SCT in the present research. In contrast, semantic priming paradigms lack the contribution of response processes, which potentially makes them less robust and more difficult to detect. This could explain why significant stereotype priming was observed only once (Study 3, race) when an LDT (semantic priming paradigm) was used in the present research. The pattern of more robust priming in response priming paradigms versus semantic priming paradigms has been established for evaluative priming, but has not been directly tested for stereotype priming (Wentura & Degner, 2010; although see Banaji & Hardin, 1996 for an indirect test). The present results, therefore, affirm the assertion that response priming paradigms give rise to more robust stereotype priming effects than semantic priming paradigms.

One immediate implication of this research concerns appropriate interpretation of stereotype priming results. Much of the research on stereotype priming has been devoted to testing the “automaticity” of stereotype activation (e.g., Blair, 2002). The distinction between semantic and response priming paradigms speaks directly to this issue. If priming effects are driven by memory activation (as in semantic priming paradigms), then priming is thought to operate at an early stage when a target concept is accessed from memory. This more directly captures what researchers refer to as “automatic activation.” Alternatively, if priming effects are driven largely by response compatibility (as in response priming paradigms), then priming is occurring at a later stage that is more dependent on the processing goals that people use to select and execute a response. Because both processes can occur in response priming paradigms, stereotype priming effects from response priming paradigms, like the SCT, do not necessarily provide evidence for direct associative links among social stimuli. Instead, they might only be evidence of a response strategy and therefore not optimal for the study of “automatic” stereotype activation.

There has been recognition that using a task like the LDT provides stronger evidence for automaticity than tasks like the SCT (e.g., Banaji & Hardin, 1996), but many stereotype priming studies have failed to consider the importance of participant task. For example, previous studies claim that eye gaze (Macrae, Hood, et al., 2002, Study 2), stimulus familiarity (Macrae, Mitchell, et al., 2002, Study 2),
negation training (Kawakami et al., 2000, Study 3), counterstereotype exposure/expectancy (Blair & Banaji, 1996, Studies 3 and 4), and stimulus configuration (Castelli et al., 2004, Studies 1 and 2; Macrae & Cloutier, 2009, Study 1b) all moderate stereotype activation. However, because each of these studies used a response priming paradigm, their manipulations may have affected response selection/execution as opposed to memory activation (Deutsch & Gawronski, 2009). That is not to say that these manipulations definitively do not alter activation, but the present evidence is insufficient to justify this claim due to the confounding presence of response competition processes in response priming paradigms.

Researchers who are primarily interested in automatic stereotype activation have two potential avenues to avoid these issues. The first option is to use a response priming paradigm and implement methods that separate out the effects of semantic and response priming processes. One method of doing this is supplementing RT measures with ERPs, which is where our research began. It may also be possible to statistically separate different processes using just behavioral measures (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Voss, Rothermund, Gast, & Wentura, 2013). Another option is to use a semantic priming paradigm in place of a response priming paradigm. It is important to note, however, that although using a semantic priming paradigm removes the confounding presence of response processes, priming observed in these paradigms does not necessarily reflect purely automatic processes (see Footnote 11). The findings of the present research also illustrate that priming effects in semantic priming paradigms (or at least the LDT) appear to be relatively inconsistent and fragile (see also Wentura & Degner, 2010 for review and suggestions). We used an LDT because it was the most prevalent semantic priming task in the literature, but there are other semantic priming tasks that might yield better results, such as the pronunciation task and semantic categorization. Indeed, our full-scale meta-analysis on sequential stereotype priming found that while stereotype priming failed to reach significance for the LDT, it was significant when a semantic categorization task was used (Kidder et al., in press). Priming was non-significant for pronunciation tasks in the meta-analysis, but this may be due to the small sample size available. Future research could directly investigate the extent to which these conclusions extend to other tasks used in semantic priming paradigms.

It is important to note that the inconsistencies observed in this set of studies do not imply that stereotype priming does not occur in semantic priming paradigms. As discussed previously, we elected to use the LDT because this task was the most prevalent in past research, and thus presumably led to the more robust findings. It may be that other tasks which engender more semantic processing of targets will lead to greater semantic stereotype priming. It is also possible that changing aspects of the LDT itself may lead to more robust priming. For example, Wentura and Degner (2010) suggest that nonword targets be similar to word targets to encourage greater depth of processing. We did this in Study 5 and found no evidence of priming, but a more systematic exploration of this might prove fruitful. It might also be useful to systematically vary whether the stereotype item appears as a target or prime. In all of our studies, the prime was an individual (male/female, black/white) and the target was a stereotype item. We chose this arrangement because this seems the most common situation that involves stereotype activation and use – encountering an unknown person and activating stereotypes to guide interactions and memory. The reverse, however, can also occur – reading the word “nurse” may activate “woman.” These and other ideas may provide avenues for future exploration.

Another objective of this paper is to serve as a cautionary example to guide future research on stereotypes and other constructs. We began this research with the idea of enhancing the study of stereotypes by employing event related potentials and proceeded under the assumption that stereotype priming effects from tasks such as the lexical decision task were established and relatively easy to replicate. This journey led to a greater appreciation for the importance of replication and follow-up research on studies that attempt to use more sophisticated methodologies to parse apart theoretical underpinnings of an observed effect. The research in stereotype priming, for example, has not focused much on the theoretical underpinnings of the priming that occurs. Research in other areas has demonstrated that there are different types of priming – some are due to “lower” types of perceptual processes such as stimulus form (e.g., Adelman et al., 2014) and others due to the nature and organization of memory (Schacter, Gutchess, & Kensinger, 2009). The latter category, which is the most relevant to stereotype priming, can be divided into the broad categories of semantic priming and associative priming. Semantic priming can be further divided into subtypes that involve categories (e.g., horse and pig because they are both animals), function (e.g., a broom and floor because one is used to clean the latter), and shared script/schema (e.g., wine and restaurant because former is

---

12 Studies 1–4 used non-words from the ARC database because the first study we performed (with ERPs) used ARC non-words. We wanted to avoid changing too many methodological factors across studies as we searched for an explanation of our null results, so we chose to keep these non-words in the next several studies.
included in the script of the latter; Moss, Ostrin, Tyler, & Marslen-Wilson, 1995). Research has revealed that these different semantic associations may be distinct from associative relations between items that are commonly linked in language use and thought (Moss et al., 1995; Postman & Keppel, 1970). Stereotype priming may be driven by both semantic and associative aspects. That is, stereotypes are frequently recognized as one component of social-cognitive schemas (e.g., Casper, Rothermund, & Wentura, 2010) and recent priming research has demonstrated their associative nature (Verhaeghen et al., 2011). Trying to differentiate the semantic and associative nature of stereotypes is potentially important because research suggests that semantic and associative relations may give rise to different priming effects (e.g., Carson & Burton, 2001; Hutchison, 2003; Lucas, 2000; Moss et al., 1995; Voss et al., 2013; Wiese & Schweinberger, 2008). While it may be arduous to create stereotype stimulus pairs that separate the script and associative relation components, the diffusion model analysis used by Voss and colleagues (2013) could be implemented to identify their differential impact on priming (see also Conrey et al., 2005). This might be particularly important for research on “automatic” stereotype priming, since previous research suggests that associative priming appears to reflect activation processes whereas semantic priming reflects response processes (Voss et al., 2013). Automaticity was not a primary focus in the current research, but it is our hope that this paper and its insights will help other stereotype priming researchers avoid some of the pitfalls that hampered our own endeavors.

Conclusions

The present research demonstrated that participant task is a significant moderator of the stereotype priming effect. Specifically, stereotype priming effects are more consistent and robust when participants perform a stereotype classification task (SCT) than a lexical decision task (LDT). This pattern of results can be understood as the difference between response and semantic priming paradigms. Response priming paradigms ostensibly produce priming effects influenced by both semantic and response priming processes, resulting in stronger priming effects. However, the confounding presence of response processes in these paradigms presents significant problems for those primarily interested in “automatic” stereotype activation. Using a semantic priming paradigm is an appealing option to address this confound, but may introduce additional complexities and concerns, including an increased chance of null results. Stereotype priming researchers are encouraged to carefully consider these matters when planning future studies.

Acknowledgments

 Portions of this research were performed under an appointment awarded to Katherine R. G. White from the US Department of Homeland Security (DHS) Scholarship Program, administered by the Oak Ridge Institute for Science and Education (ORISE). ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE Contract Number DE-AC05-06OR23100. All opinions expressed in this paper are the authors’ and do not necessarily reflect the policies and views of DHS, DOE, or ORAU/ORISE. We thank Adriaan Spruyt for helpful advice regarding the mental tally procedure in Study 2 and Ciara Kidder for her assistance with the meta-analytic analyses.

References

Appendix A

Primes and Target Stimuli for Studies 1–5

Study 1

Primes

Pictures of unknown males and females were gathered from a public rating forum called “Rate My Picture” on https://www.MySpace.com. Pictures were initially gathered and then screened for unusual or distracting features (e.g., green hair, busy background). The remaining pictures were rated for attractiveness by 47 individuals on a scale from −3 to +3. The present study used 54 pictures of males and 54 pictures of females as primes, all neutral in attractiveness.

Names were chosen from a list of approximately 600 male and female names that were rated by 20 individuals along the dimensions of masculinity/femininity and familiarity. The masculinity/femininity of names was rated on a scale from −3 (very masculine) to +3 (very feminine). Familiarity was rated on a scale from −3 (very unfamiliar) to +3 (very familiar). The 54 most masculine and familiar names were chosen to serve as male primes, and the 54 most feminine and familiar female names were chosen to serve as female primes. The average masculinity rating for male names was −2.55, with a range of −2.2 to −2.95. The average familiarity rating for male names was 1.69, with a range of 1–2.76. For female names, the average femininity rating was 2.62, with a range of 2.4–2.95. The average familiarity rating for female names was 1.63, with a range of 1–2.76.

Male Names: JACOB, RICARDO, MARCOS, RAUL, CHRISTOPHER, JESUS, JOSEPH, ROBERT, ANTONIO, ARTURO, EDUARDO, GEORGE, HUGO, MANUEL, MARK, MATTHEW, PETER, ROBERTO, ALFREDO, DAVID, HECTOR, PABLO, PAUL, PEDRO,
Gender stereotypical words were obtained in a series of steps. First, 18 undergraduate students were asked to list personality traits, occupations, and objects stereotypically associated with men and women. Second, these were entered into a database and the most frequently generated words were identified. Lastly, these 70 words were rated by an independent set of 44 undergraduate students for masculinity/femininity (1 = very masculine, 7 = very feminine) and valence (−3 to +3). The 8 most feminine traits (M = 5.45) and 8 most feminine nontraits (M = 6.23; e.g., occupations, objects, activities) were chosen to serve as female stereotypical targets. The 8 most masculine traits (M = 2.34) and 8 most masculine nontraits (M = 1.66) were chosen to serve as male stereotypical targets. Male and female targets were significantly different in their masculinity/femininity ratings, t = 19.88, p < .002, and were matched for valence, p = .404, length, p = .544, and frequency, p = .287.

Female Traits: GOSSIPY, NURTURING, EMOTIONAL, DELICATE, TALKATIVE, SENSITIVE, CARING, VULNERABLE
Female Nontraits: DRAMA, LIPSTICK, PURSE, MANICURE, ESTROGEN, NURSE, SECRETARY, BEAUTICIAN

Male Traits: MACHO, TOUCH, ROUGH, VIOLENT, COCKY, STRONG, AGGRESSIVE, DOMINANT
Male Nontraits: TESTOSTERONE, TRUCKS, PLUMBER, FOOTBALL, SOLDIER, CONSTRUCTION, MECHANIC, MUSCLES

Nonwords: WRYNKED, KREMBTH, EEBS, STRYLTH, PEUNTHS, SKWUXTE, SWIETH, PHROGNS, SPRUPE, SPRIRGUE, SPOMFS, FROUHGHE, PRAFES, CWERSH, SWALS, SHRUPES, SPENCHED, TWILS, GLAUGHTH, GHUGNTH, SPLERLTE, TROUZE, ZYMPH, BLEIGG, REGUED, SNOLFS, GHWUCSTS, PSANCSED, THROOMS, WEUSED, PLAUGHGUES, GHWOURV, CURDGED, GWURP, SKIRCHED, SPRURVES, TWOMF, DWEAMTHS, SCKERGNS, SKWYLMB, TRAUHGED, SOCH, GHEILTS, ZEIDGE, LURCED, BONTHS, SNOURMB, JERG, GNOWLE, WERCHRED, CRINSE, FROINED, WOWGE, CRIGHV, GEELLS, PRAUGHD

Study 2
Same as Study 1.

Study 3

Primes
Pictures of unknown males and females were gathered at the University of Texas at El Paso and used courtesy of Michael Zárate. They were head and shoulder pictures of individuals on a neutral background.

Targets
Race stereotypical words were obtained in a series of steps. First, we identified a set of items stereotypically associated with blacks, whites, and Hispanics. Some stereotype terms were found in published research, and we added to this set by asking approximately 10 individuals in El Paso, Texas (primarily undergraduate students) to generate stereotype words for blacks, whites, and Hispanics. This resulted in a list of 270 potential items. Second, we then had 42 undergraduate students rate these 270 items. Half rated them on a 7-point Hispanic/black dimension and a 7-point valence dimension; and half rated the items on a 7-point white/black dimension and a 7-point valence dimension. We then elected to focus the study on black/Hispanic stereotypes and selected the strongest set of 20 stereotype items for each group that were closely matched in stereotypicality (M = 1.94 vs. 6.08 for Hispanic and black terms, respectively, on a scale with 1 = Hispanic, 4 = Race Neutral, 7 = Black) and valence (M = 4.59 and 4.56 for Hispanic and black terms, respectively, on a scale with 1 = Negative, 4 = Neutral, 7 = Positive).

Hispanic: CHILI, DAY LABORER, MAIDS, FAMILY, MACHO, SHORT, TRADITION, ENCHILADAS, LATIN AMERICA, MARIACHI, TEQUILA, SOCCER, CUMBIA, DEPORTATION, IMMIGRANT, SAMBA, TANGO, ACCENT, CONSTRUCTION, CATHOLICS
Black: SOUL, ATHLETE, RAPPER, BLACK PANTHERS, SWAGGER, TALL, SPORTS, FOOTBALL, CHURCH CHOIR, RUNNER, LOUD, JAZZ, R&B, SEGREGATION,
SLAVERY, HIP HOP, BASKETBALL, AFRICA, NFL, BLING


Study 4

Primes

Same as Study 3.

Targets

The Hispanic & black stereotype words and nonwords were the same as Study 3. The process for identifying gender stereotype words was identical to that described in Study 3, but the original files and data from the pilot study could not be located so we are unable to report mean stereotypicality and valence for the male and female stereotype words. The words used in the study are in the table below.


Race

Sixteen participants provided valence (1 = very negative, 7 = very positive) and race association (1 = very associated with Blacks, 7 = very associated with Whites) ratings for the race target stimuli. The 30 words most with Blacks and 30 words most associated with Whites were then selected for use in the study. These stimuli were significantly different regarding their race association ratings (M_B = 2.32, M_W = 5.73), t = –37.21, p < .001, but were matched for valence (M_B = 3.44, M_W = 3.95), t = –1.66, p = .102. Nonwords were created by changing one or two letters in the original target word to create a pronounceable nonword.

Blacks: rappers, ghetto, dreads, weaves, basketball, gangs, rims, streetwise, foodstamps, poverty, rhythm, soulful, welfare, violence, mistreated, watermelon, dancing, unemployed, uneducated, prison, projects, collards, chicken, struggle, fighting, jazz, lips, reunions, ribs, dangerous

Whites: country, beach, uptight, snobby, lawyers, power, pink, politics, swimming, boss, suburbs, trailers, superiority, patriotic, shotgun, honkie, Abercrombie, farming, presidents, majority, skiing, republican, pale, hunting, privileged, racist, golf, tennis, tanning, redneck

Race Nonwords: roopers, guegto, drefs, woales, baspethall, golgs, riws, straectcie, foolslamps, pazerty,
rhuhm, syllkul, wilfire, vielesce, mustgeated, witertel-
lon, domcing, unimzloyed, unewudated, plicon, prokacts,
coltords, choken, steohgle, feghking, jal, lups, reinisns,
rebs, dangproos, equttry, boach, uppight, sbobery, laquers,
puwtr, ponk, poftids, swomling, byss, sipurbes, treifers,
sapertority, perioci, shuttun, hyngie, azerqrombie, ferp-
ing, preshebts, mipority, swuing, redulican, paie, hon-
jing, pyvileweed, rikist, guff, ternas, talping, reffeck

Appendix B

Data Preparation for Studies 1-5

Except as noted below, we used an identical procedure to
to screen and prepare data for analyses across all five studies.
First, responses within 250 ms of target onset were elimi-
nated. Second, we eliminated data from participants whose
overall accuracy was more than two standard deviations
below the mean accuracy for their between-subject condi-
tion (if the experiment had a between-subject manipula-
tion). Third, response times (RTs) greater than 2 standard
deviations from each participant’s mean response latency
were replaced by the mean plus 2 SD values for that partic-
ipant (see Wittenbrink, 2007 for similar criterion). Fourth,
we eliminated data from participants if one or more
within-subject conditions contained less than 1/3 of avail-
able trials. This is done to ensure that each within-subject
condition had enough trials to obtain a stable average for
that condition. Participants sometimes have a response bias
(e.g., respond “nonword” in LDT, or “male” in SCT) that
results in fewer trials in certain conditions, but the bias is
not strong enough to lower their overall accuracy rate
below the 2 SD accuracy threshold for excluding all of their
data. After these steps, response times from correct
responses were averaged in each condition. In LDT condi-
tions, the nonword trials were not examined.

In addition to the analyses reported in the text, we did
two other sets of analyses on the data in all five studies
to support the results reported in the text. First, to ascertain
whether steps were followed to exclude data impacted find-
ings, we analyzed the data from all participants (including
those who had low accuracy or were missing more than
2/3 of available trials in a within-subject condition). The
pattern of results and significance of these analyses are
identical to those reported in the primary analyses for all
five studies, save for one finding in Study 3 that was
significant in the primary analyses and marginal in the
analysis with all participants (see Footnote 10). Second,
we examined both the untransformed RT data and also
RT data that was subjected to an inverse transform. The
pattern of results and significance for these analyses were
identical across all five studies. We report the analyses on
untransformed RT data for ease of interpretation.

Reported below are the numbers associated with these
data preparation steps for each of the five studies.

Study 1

- Data from 11 participants were not available because of
technical problems with hardware/software during the
experiment.
- Data from 12 participants excluded due to low accuracy:
  Four participants in the LDT condition (M = 94.7%;
  SD = 5.0%); three in the pre-primed LDT condition
  (M = 94.6%; SD = 5.7%), and five in the gender catego-
  rization condition (M = 90.9%; SD = 11.2%).
- Data from five participants excluded due to low number
  of trials in one or more conditions (1 in LDT, 1 in pre-
  primed LDT, and 3 in gender categorization).
- For the participants included in the primary analyses,
  5.3% of trials were eliminated because response
  occurred within 250 ms of target onset. Although we
  begin data preparation by removing fast responses, we
  report the percentage of fast response only for partici-
  pants included in the analyses. A handful of participants
  engage in a practice of responding quickly (and ran-
  domly) to targets so they can quickly finish the experi-
  ment (accuracy for fast responses is approximately
  chance across all participants). Thus, including partici-
  pants who are removed due to low accuracy inflates
  percentage for participants whose data is analyzed.

Study 2

- Data from two participants were not available because of
technical problems with hardware/software during the
experiment.
- Data from 11 participants excluded due to low accuracy:
  Eight participants in the ethnicity tally condition
  (M = 87.9%; SD = 11.2%) and three in the gender tally
  condition (M = 89.1%; SD = 8.8%).
- Data from 15 participants excluded due to low number
  of trials in one or more conditions (11 in ethnicity tally
  and 4 in the gender tally).
- For the participants included in the primary analyses,
  13.7% of trials were eliminated because response
  occurred within 250 ms of target onset.
Study 3

- Data from three participants were not available because of technical problems with hardware/software during the experiment.
- Data from 22 participants excluded due to low accuracy: Five participants in the LDT condition ($M = 85.3\%; SD = 14.4\%$) and 17 in the SCT condition ($M = 77.1\%; SD = 18.0\%$). The cutoff number used for elimination in the SCT condition is one change from the typical procedure. Because of the low mean accuracy and high standard deviation in this condition, the two standard deviation cutoff fell below chance levels of responding. So, we used the same cutoff in the SCT condition as used in the LDT condition (65.5%).
- Data from six participants excluded due to low number of trials in one or more conditions (2 in SCT and 4 in LDT).
- For the participants included in the primary analyses, 8.7% of trials were eliminated because response occurred within 250 ms of target onset.

Study 4

- Data from 11 participants excluded due to low accuracy: nine in the gender-LDT condition ($M = 87.9\%; SD = 13.8\%$) and two in the race-LDT condition ($M = 91.3\%; SD = 8.3\%$).
- Data from two participants excluded due to low number of trials in one or more conditions (1 in gender task and 1 in race task).
- For the participants included in the primary analyses, 0.9% of trials were eliminated because response occurred within 250 ms of target onset.

Study 5

- Data from 20 participants excluded due to low accuracy: thirteen in the race-SCT condition ($M = 78.7\%; SD = 16.5\%$); two in the gender-SCT condition ($M = 91.5\%; SD = 9.0\%$); one in the race-LDT condition ($M = 93.4\%; SD = 6.9\%$); four in the gender-LDT condition ($M = 91.7\%; SD = 10.7\%$).
- Data from one participant (gender-LDT condition) excluded due to low number of trials in one or more conditions.
- For the participants included in the primary analyses, 0.9% of trials were eliminated because response occurred within 250 ms of target onset.