Stereotype Validation: The Effects of Activating Negative Stereotypes After Intellectual Performance

Jason K. Clark and Kelsey C. Thiem
University of Iowa

Jamie Barden
Howard University

Jillian O'Rourke Stuart
University of Iowa

Abigail T. Evans
Ohio State University

With regard to intellectual performance, a large body of research has shown that stigmatized group members may perform more poorly when negative, self-relevant stereotypes become activated prior to a task. However, no research to date has identified the potential ramifications of stereotype activation that happens after—rather than before—a person has finished performing. Six studies examined how postperformance stereotype salience may increase the certainty individuals have in evaluations of their own performance. In the current research, the accessibility of gender or racial stereotypes was manipulated after participants completed either a difficult math test (Studies 1–5) or a test of child-care knowledge (Study 6). Consistent with predictions, stereotype activation was found to increase the certainty that women (Studies 1, 2, 4, and 5), African Americans (Study 3), and men (Study 6) had toward negative evaluations of their own test performance. These effects emerged when performance-related perceptions were stereotype consistent rather than inconsistent (Studies 1–6) and were found to be most pronounced among those who were highly identified with the stereotyped group (Study 5). Furthermore, greater certainty—triggered by negative stereotypes—predicted lowered domain-relevant beliefs (Studies 1, 2, 3, and 6) and differential exposure to domain-relevant stimuli (Studies 4 and 5).

Keywords: certainty, metacognition, evaluations, stereotypes, test performance

A wealth of research on stereotype threat has shown that intellectual performance can be inhibited when self-relevant, negative stereotypes become accessible (see Steele, Spencer, & Aronson, 2002). Although the extant literature has focused almost exclusively on stereotype awareness that occurs *prior* to a task, there may be many situations in which negative stereotypes only become salient *after* an individual has finished performing. Consider a group of stigmatized individuals who report their gender and ethnicity after they complete an intellectual test. Field research suggests that underperformance indicative of stereotype threat might be reduced substantially—by as much as 33%—when simply answering group-identifying questions after rather than before the test (Danaher & Crandall, 2008; but see Stricker & Ward, 2004). As some have advocated, moving demographic inquiries to the end of standardized tests may be one of the most straightforward, cost-effective ways to lessen group-based performance gaps (e.g., see Danaher & Crandall, 2008; Jordan & Lovett, 2007). However, does this sequence hold the potential to affect individuals in a new way—one that is distinct from stereotype threat and other identified influences on self-perceptions? Building from recent advances in the study of metacognition (Clark, Wegener, Briñol, & Petty, 2009), the present research examined the possibility that negative stereotypes can affect the certainty that individuals have in their performance-related perceptions.

Metacognition, Self-Validation, and Stereotypes

Metacognition refers to people's perceptions about their own thoughts or thought processes (i.e., cognition about cognition; see Jost, Kruglanski, & Nelson, 1998). One metacognitive aspect of thinking that has received substantial attention is the degree to which people are certain of their thoughts and attitudes about information in their social environment. According to the self-validation hypothesis (Petty, Briñol, & Tormala, 2002), the degree of certainty that one has in their thoughts should have a significant impact on the nature of the relevant attitudes or evaluations that are formed. For example, two people could produce the same negative thought toward a stimulus, but one person could be more certain of this reaction

Jason K. Clark and Kelsey C. Thiem, Department of Psychology, University of Iowa; Jamie Barden, Department of Psychology, Howard University; Jillian O'Rourke Stuart, Department of Psychology, University of Iowa; Abigail T. Evans, Department of Psychology, Ohio State University.

This research was supported in part by National Science Foundation Grants 1226417 and 1225804. We are grateful for the helpful comments and contributions provided by Catherine Drahota, Meara Habashi, Molly McLeese, Jacob Sullivan, and Andrew Todd.

Correspondence concerning this article should be addressed to Jason K. Clark, Department of Psychology, University of Iowa, Iowa City, IA 52242. E-mail: jason-k-clark@uiowa.edu

Journal of Personality and Social Psychology, 2015, Vol. 108, No. 4, 531–552
 0022-3514/15/$12.00 http://dx.doi.org/10.1037/a0038887

than the other individual. As a result, the person who is more certain of their negative thought should form an evaluation that is also held with greater certainty (see Petty et al., 2002).

Based on a wealth of previous research, this key difference in certainty should have critical implications for how durable the evaluation is and how much influence it may have on related beliefs, judgments, and behaviors (for reviews, see Petty, Briñol, Tormala, & Wegener, 2007; Tormala & Rucker, 2007). For instance, research has found that evaluations held with high certainty are more difficult to change (e.g., Bassili, 1996), exert greater bias on related beliefs (e.g., Marks & Miller, 1985), and guide behavior to a greater extent (e.g., Fazio & Zanna, 1978) compared with evaluations held with less certainty. A number of factors have been shown to induce certainty; including the belief that one shares consensus views on an issue (e.g., Petty et al., 2002), believing that a source of information is highly credible (e.g., Tormala, Briñol, & Petty, 2007; cf. Clark & Evans, 2014), and experiencing positive mood (Briñol, Petty, & Barden, 2007). Moreover, research has shown that many of these variables evoke certainty often when they are introduced *after* rather than before one has produced thoughts in response to a stimulus (see Briñol & Petty, 2009).

Although these processes have received considerable attention in research on attitudes and persuasion, few investigations have explored metacognitive validation in other domains of social and self-evaluation. However, some research has examined how social stereotypes may elicit certainty in the perceptions that perceivers form of other individuals (Clark et al., 2009). Stereotypes are cognitive representations of several aspects of group memberships, including physical features, traits, and behaviors. These representations have been shown to guide perceptions in many distinct ways (for a review, see Bodenhausen, Macrae, & Sherman, 1999). It stands to reason that stereotypes could also serve to confirm a perception—and therefore increase the certainty one has toward it. In a set of experiments, Clark et al. (2009) found increased certainty when the nature of one's perceptions was consistent with or matched a *later* activated stereotype about another person. For instance, after thinking that a target person was unintelligent, participants were more certain and expressed stronger perception-consistent judgments when they later learned that this person was low as opposed to high in socioeconomic status (SES).

Clark et al. (2009) reasoned that this certainty effect emerged in a way that is similar to how convergent validity is obtained in science (Campbell & Fiske, 1959). In particular, acquiring information from multiple sources (e.g., perceptions about a target's attributes and a later activated stereotype) that supported or converged with one another (e.g., thoughts about unintelligence and low SES) resulted in greater perceptual certainty (see also Clark, Wegener, Sawicki, Petty, & Briñol, 2013). Moreover, this conceptualization aligns with several other past investigations of evaluative certainty (for a review, see Rucker, Tormala, Petty, & Briñol, 2014). Most notably, research has found that people report greater certainty when they learn that their opinion on an issue represents a consensus viewpoint (e.g., Petrocelli, Tormala, & Rucker, 2007) and when acquired information about a stimulus is consistent rather than mixed in terms of valence (Smith, Fabrigar, MacDougall, & Wiesenthal, 2008).

Implications for Self-Relevant Perceptions

Beyond impressions of others, it is plausible that validation effects could also occur when perceptions are about the self and stereotypes are self-relevant. Over the past two decades, some of the most widely studied stereotyping phenomena correspond to effects of stereotype threat. *Stereotype threat* has been characterized as a psychological predicament in which individuals fear that their performance may be evaluated in light of a negative stereotype of their group (see Steele, 1997). In these situations, the possibility of confirming a stereotype is thought to place added pressure that can subsequently undermine an individual's performance. In a seminal set of studies, Steele and Aronson (1995) found that Blacks performed worse than Whites when stereotypes about intellectual ability were activated prior to taking a test (e.g., reporting one's race). However, when the stereotype was less salient (e.g., not reporting race before the test), the performance of Black and White students was equivalent when controlling for previous scores on the Scholastic Aptitude Test (SAT). Since the publication of this work, research on stereotype threat has been extended in a number of different directions, such as identifying other at-risk group memberships (e.g., women in math, Spencer, Steele, & Quinn, 1999), underlying mechanisms (e.g., decreased working memory, Schmader & Johns, 2003), and intervention strategies (e.g., affirming self-worth, Martens, Johns, Greenberg, & Schimel, 2006).

With regard to the latter, one straightforward practice to combat stereotype threat may be to simply move demographic questions of gender and ethnicity from the beginning to the end of standardized tests. In their reanalysis of a field study on the Advanced Placement Calculus exam (Stricker & Ward, 2004), Danaher and Crandall (2008) asserted that women tended to perform better when demographics were answered after rather than before the primary test questions. Although this simple change in testing procedures has been estimated to produce a substantial reduction in performance bias (see Danaher & Crandall, 2008), it could ultimately produce another kind of harmful effect. Clearly, posttest stereotype activation cannot influence how a person performed. However, it may impact how one *perceives* their own performance—in particular, by validating negative perceptions.

Consistent with the aforementioned convergence account proposed by Clark et al. (2009), this potential for *stereotype validation* may be most likely when an activated stereotype matches the nature of the perceptions that one has produced. Consider a group of males and females who take a very difficult math test—wherein the salience of gender stereotypes (and likelihood of stereotype threat) has been limited prior to the task. Because the test is challenging, a majority of the test takers should produce largely negative perceptions (Cadinu, Maass, Rosabianca, & Kiesner, 2005; Schmader, Forbes, Zhang, & Mendes, 2009) and believe that they performed poorly on the test. Following the test, stereotypes are now made accessible by asking the test takers to report their gender on a demographic question. For women who viewed their performance as poor, the female stereotype (women are bad at math) is consistent with this negative evaluation and the difficulty they likely experienced while taking the test. Thus, because of this convergence between perceptions and the later activated stereotype, women should feel more certain of their negative performance-related perceptions. For men who believe their performance was poor, making their gender salient should not yield the same effect. In this case, the activated stereotype (men are good at math) is not consistent with their test experience and, therefore, should not increase certainty in their negative performance-related perceptions.

These postulated effects should have important strength-related implications. As previously described, a large body of research has shown that evaluations held with greater certainty are stronger and more effectively guide related beliefs and behaviors compared with evaluations held with less certainty (for reviews, see Petty & Krosnick, 1995; Tormala & Rucker, 2007). Therefore, it stands to reason that feeling more certain that one has performed poorly may yield a number of harmful downstream consequences. For instance, in the previous example, women who were validated by the gender stereotype may show decreased beliefs in their own mathematical abilities, altered academic/career interests, and may be more likely to behave in stereotype-consistent ways in math-relevant contexts.

Research Overview

The present research examined how activating negative stereotypes after performance may increase the certainty individuals have in their negative performance-related evaluations. We predicted that this effect should be most likely to occur when an individual forms a negative evaluation of their own performance and a later activated stereotype is consistent with this appraisal. These predictions were tested across six studies. In each study, participants completed a set of difficult test questions. After the test, the accessibility of certain group performance stereotypes was manipulated and participants subsequently completed dependent measures corresponding to their performance and other domain-relevant perceptions. Study 1 examined gender stereotypes in the context of math performance and served as an initial test of the proposed stereotype validation effect. Furthermore, this study examined the necessity of a match between a stereotype and perceived performance. The aim of Study 2 was to build upon the previous findings by using a different manipulation of stereotype accessibility and measuring downstream outcomes such as math-related career interests and expectations of future performance.

Following the method of Study 2, we examined stereotype validation with regard to math performance and stereotypes of African Americans in Study 3. Studies 4 and 5 once again focused on the math performance of women and examined how stereotype validation may have implications for future behavior. In particular, we investigated how this phenomenon may guide choices regarding the types of math-related information that individuals seek out. In addition, Study 5 examined the extent of self-identification with a negatively stereotyped group as a potential moderator of these effects. Lastly, the aim of Study 6 was to investigate the implications that stereotype validation may hold for other performance domains beyond math. Specifically, we examined the potential for stereotype validation in the context of gender stereotypes about child-care abilities.

Study 1

In Study 1, male and female college participants completed a challenging math test. After finishing the test and rating their perceived performance (either *poor* or *strong*), participants either reported their gender on a demographic question or received no such inquiry. Given the difficulty of the test, we expected a large majority of the participants to believe they had performed poorly on the test. However, the accessibility of gender should not influence each group the same way. Because the female stereotype (women are bad at math) converges with their performance experience (difficulty), women who believed they had performed poorly should feel more certain of their negative perceptions when gender is made salient after the test compared with the control. However, for men who viewed their performance as poor, the stereotype (men are good at math) does not converge with the difficulty they experienced while taking the test. Thus, activating this aspect of their identity should do little to influence the certainty they have in their performance-related perceptions. These hypothesized differences in evaluative certainty may carry important consequences. With this in mind, participants in Study 1 also reported their beliefs in their own math skills and abilities near the conclusion of the study. We predicted that greater certainty, triggered by the gender stereotype, should predict lowered beliefs in one's math abilities.

Method

Participants and design. Two hundred eighty-two undergraduates (132 women and 150 men) at a large Midwestern U.S. university participated for partial course credit. Approximately 80% of participants (234 of 282) were White. Participants were randomly assigned to a condition in which their gender was or was not made salient after completing a set of math problems. The study represented a 2 (gender: women, men) $\times$ 2 (performance evaluation: poor, strong) $\times$ 2 (gender demographic inquiry: absent, present) between-participants design.

Procedure. A maximum of three participants of the same gender took part in any single session. Upon arrival to the lab, each participant was greeted by an experimenter of the same gender and was seated at a private computer station that was visually isolated from others. In an attempt to further limit the likelihood of stereotype threat, participants were told this was a problem-solving exercise and no reference to gender was made (instructions adapted from Rydell, McConnell, & Beilock, 2009). All participants then completed 12 difficult math problems drawn from practice GRE tests (approximately 50% accuracy from past examinees; see Educational Testing Service, 1998). Participants were given an unlimited amount of time and were required to answer each question.

Immediately after the test, participants rated their own performance on a dichotomous measure (*poor* or *strong*) and responses were used to index each participant's performance evaluation. The manipulation of gender salience followed this rating. Participants assigned to the gender inquiry-present condition reported their gender by clicking one of two boxes—either *male* or *female*. In contrast, participants assigned to the gender inquiry-absent condition were not asked to report their gender. After this induction, participants completed scaled measures of evaluative certainty, perceived performance, and math ability beliefs. Finally, participants reported their race and were debriefed.

Dependent measures.

Math performance. All participants completed the same number of math questions (12) and the total number of correct answers served as the index of test performance.

Evaluative certainty. Following the gender-salience manipulation, participants reported the certainty they had in their performance evaluation on eight 11-point scales. The first four items matched each participant's response on the dichotomous performance evaluation. Participants who rated their performance as "poor" on this measure received the following items: "How certain are you that your performance on the test was poor?" (1 = *not at all certain* to 11 = *very certain*); "How sure are you that your performance on the test was poor?" (1 = *not at all sure* to 11 = *very sure*); "I am certain that I performed poorly on the test" (1 = *strongly disagree* to 11 = *strongly agree*); and "I am sure that I performed poorly on the test" (1 = *strongly disagree* to 11 = *strongly agree*). Participants who rated their performance as "strong" received similar questions, but the word *poor/poorly* was replaced with *strong/strongly*. The remaining four measures were not performance-specific and the wording of the questions was the same for all participants. These items read as follows: "To what extent are you certain that your perceptions of your performance are accurate?" (1 = *not at all certain* to 11 = *very certain*); "To what extent are you sure that your perceptions of your performance are accurate?" (1 = *not at all sure* to 11 = *very sure*); "In general, how accurate do you think your perceptions of your performance are?" (1 = *not at all accurate* to 11 = *very accurate*); and "My perceptions of how I performed on the test are accurate" (1 = *strongly disagree* to 11 = *strongly agree*). Responses to these measures were averaged to form a composite ($\alpha = .91$).

Perceived math performance. After the certainty measures, participants rated their math performance on two 11-point scales: "Overall, how well do you think you performed on the math problems?" (1 = *performed extremely poorly* to 11 = *performed extremely well*) and "Which choice best reflects the amount of questions that you believe that you answered correctly?" (1 = *a very small amount* to 11 = *a very large amount*). Responses were averaged to form a single index ($\alpha = .94$).

Math ability beliefs. Participants reported beliefs about their math ability on four 11-point scales. The items were "Please rate your own math skills on the following scale." (1 = *very low* to 11 = *very high*); "To what extent do you believe that your math skills need improvement?" (1 = *not at all* to 11 = *very much*; reverse-scored); "Compared to other students, my math ability and skills are weak" (1 = *strongly disagree* to 11 = *strongly agree*; reverse-scored); and "Compared to other students, my math ability and skills are strong (1 = *strongly disagree* to 11 = *strongly agree*)." A single index was created by averaging responses to these measures ($\alpha = .88$).

Results

Perceived math performance. On the dichotomous performance evaluation, approximately 65% of participants reported "poor" performance (86 females and 94 males) and 35% rated their performance as "strong" (46 females and 56 males). The index formed from responses to the 11-point items was submitted to a 2 (gender: male, female) × 2 (performance evaluation: poor, strong) × 2 (gender demographic inquiry: absent, present) analysis of variance (ANOVA). This analysis revealed a main effect of the dichotomous performance evaluation, ($M_{poor} = 4.20$ [$SD = 1.61$] vs. $M_{strong} = 7.64 =$ [$SD = 1.70$]), $F(1, 274) = 282.12$, $p < .001$, $r = .71$. However, no other main effects or interactions were significant ($ps > .16$).

Actual math performance. A three-way ANOVA was also conducted on the number of correct test answers. Results showed a main effect of the performance evaluation ($M_{poor} = 2.93$ [$SD = 1.91$] vs. $M_{strong} = 5.63$ [$SD = 2.95$]), $F(1, 274) = 84.99$, $p < .001$, $r = .49$. In addition, a main effect of gender was also found, $F(1, 274) = 7.13$, $p = .008$, $r = .16$. As in previous research that has used difficult math tests (e.g., Spencer et al., 1999), male participants ($M = 4.27$, $SD = 2.67$) performed better than female participants ($M = 3.49$, $SD = 2.61$) on the problem set. No other main effects or interactions were found ($ps > .53$).

Evaluative certainty. A three-way analysis of covariance (ANCOVA) that controlled for the number of correct test answers (i.e., actual math performance; $M = 3.90$, $SD = 2.67$) was conducted on the index of evaluative certainty. The results revealed a pattern of effects that was consistent with the primary hypotheses. In particular, a significant Gender × Performance Evaluation × Gender Inquiry interaction was found, $F(1, 273) = 4.93$, $p = .027$, $r = .13$ (see Figure 1).[1] Among female participants, the data supported the posited convergence rationale. Women who perceived their performance to be poor were more certain of this evaluation when gender was made salient (adjusted $M = 7.28$, $SE = .28$) compared with when it was not (adjusted $M = 6.42$, $SE = .26$), $F(1, 273) = 4.94$, $p = .027$, $r = .13$. Conversely, consistent with a lack of stereotype convergence, no evidence of validation was found for women who perceived strong performance (adjusted $M_{present} = 6.67$ [$SE = .35$] vs. adjusted $M_{absent} = 7.14$ [$SE = .39$]), $F < 1$; Performance Evaluation × Gender Inquiry, $F(1, 273) = 4.65$, $p = .032$, $r = .13$.

For male participants, the pattern of results was considerably different. The gender demographic manipulation had no influence on the certainty of men who evaluated their performance as poor (adjusted $M_{present} = 6.77$ [$SE = .25$] vs. adjusted $M_{absent} = 6.73$ [$SE = .26$]), $F < 1$. Likewise, this manipulation did not have a significant effect among men who perceived their performance as strong (adjusted $M_{present} = 7.44$ [$SE = .38$] vs. adjusted $M_{absent} = 6.81$ [$SE = .32$]), $F(1, 273) = 1.52$, $p = .218$; Performance Evaluation × Gender Inquiry, $F < 1$. Beyond the three-way interaction, the only other influence that emerged was a marginally significant main effect of the actual performance covariate, $F(1, 273) = 2.77$, $p = .097$ (all remaining $ps > .23$).[2]

[1] A three-way ANOVA that did not control for the number of correct test answers revealed a similar pattern of results. In particular, the Gender × Performance Evaluation × Gender Inquiry interaction was significant, $F(1, 274) = 4.85$, $p = .029$, $r = .13$.

[2] A four-way ANCOVA was also performed that included participant race (White vs. non-White) as an independent variable. This analysis revealed no significant main effects or interactions of the race of participants ($ps > .09$). Furthermore, this same race classification was treated as an independent variable in supplementary analyses of evaluative certainty in Studies 2, 4, 5, and 6. No main effects or interactions of participant race emerged from any of these tests ($ps > .14$).

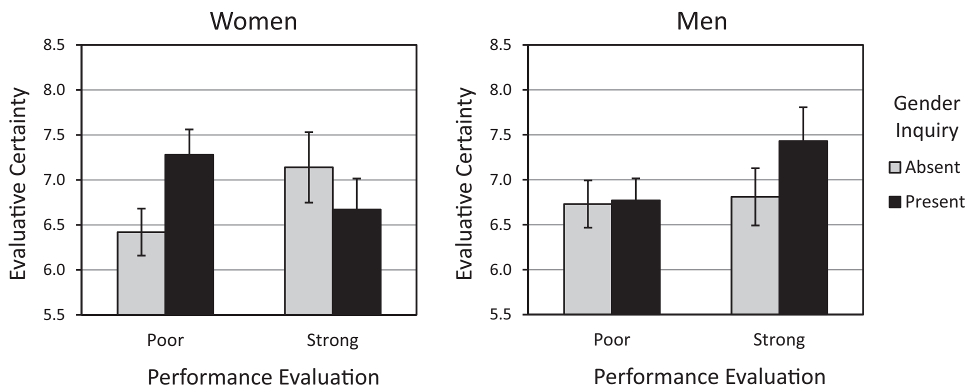

Figure 1. Adjusted mean evaluative certainty in Study 1 as a function of participant gender, performance evaluation, and the gender inquiry manipulation (controlling for actual performance on the math test).

Direct and indirect effects on math ability beliefs. A three-way ANCOVA that controlled for actual math performance was conducted on the ability beliefs composite. A significant effect of performance evaluation emerged such that beliefs were lower among participants who rated their performance as poor (adjusted $M = 6.00$, $SE = .14$) compared with strong (adjusted $M = 7.59$, $SE = .20$), $F(1, 273) = 37.78$, $p < .001$, $r = .35$. In addition, a marginally significant main effect of gender was found, $F(1, 273) = 3.08$, $p = .081$. Male participants (adjusted $M = 7.00$, $SE = .16$) tended to report higher beliefs than female participants (adjusted $M = 6.60$, $SE = .17$). Lastly, the covariate of actual performance on the test carried a significant effect, $F(1, 273) = 17.65$, $p < .001$, $r = .25$. No other effects approached significance (ps $> .15$).

As previously discussed, a broad base of previous research indicates that greater certainty should result in evaluations that are stronger and more likely to affect corresponding beliefs and behaviors (see Petty & Krosnick, 1995). In the current study, women who believed they had performed poorly on the math task were more certain of this evaluation when they later reported their gender compared with when they did not. We postulated that this enhanced certainty about poor performance should have a negative effect on relevant beliefs about one's math skills and ability. The direct effect of the manipulation on ability beliefs was found to be weak in the previously described ANCOVA. However, tests of mediation remain appropriate without a significant relationship between the independent and dependent variable (see MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Rucker, Preacher, Tormala, & Petty, 2011; Shrout & Bolger, 2002). In sum, for women who saw their performance as poor, we postulated that the gender inquiry manipulation should have an indirect effect on math ability beliefs via differences in evaluative certainty.

Following recommendations by Preacher, Rucker, and Hayes (2007; see also Hayes, 2013), multiple regression was used in conjunction with bootstrapping procedures to identify the circumstances in which an indirect effect of the manipulation may have occurred. The analysis used the PROCESS macro developed by Hayes (2014) and estimated the indirect effect for women and men at both poor and strong performance evaluations (respectively)—while also controlling for actual math test performance. In line with the previously described ANCOVA results (i.e., a significant Gender × Performance Evaluation × Gender Inquiry interaction), this analysis allowed the impact of the gender inquiry manipulation on evaluative certainty to depend on participants' gender and their performance evaluation (poor or strong perceived performance). At both poor and strong perceived performance (respectively), certainty was predicted to have the same influence on ability beliefs regardless of gender. Hence, the analysis allowed the association between evaluative certainty and math ability beliefs to depend on the performance evaluation, but not participant gender. Furthermore, PROCESS Model 68 was used; the obtained data were treated as the population; and 10,000 bootstrap samples were drawn (with replacement) to produce 95% bias-corrected confidence intervals (BC CIs).

The results of this analysis are displayed in Table 1 (see Parade, Leerkes, & Blankson, 2010, for a similar approach). As previously described, the gender inquiry manipulation significantly influenced certainty for women who evaluated their performance as poor. In turn, higher levels of certainty among these women significantly predicted decreased beliefs in one's math ability. This indirect effect of the gender inquiry manipulation was found to be significant (i.e., 95% BC CI did not include 0). Very different patterns emerged among the other groups of participants. For men who viewed their performance as poor, higher evaluative certainty predicted lower math ability beliefs. On the other hand, greater certainty tended to predict enhanced ability beliefs for both women and men who rated their test performance as strong. However, for each of these three groups, differences in evaluative certainty did not mediate the relation between the gender inquiry manipulation and math ability beliefs (i.e., 95% BC CIs for the indirect effects included 0).

Discussion

Study 1 provided initial evidence that negative stereotypes activated after math performance can validate perceptions and, in turn, may negatively affect beliefs about one's math ability. In particular, the findings suggest that this effect can occur when a

Table 1
Results of Bootstrapping Moderated Mediation Analyses in Study 1

Variable	Gender inquiry manipulation → Evaluative certainty		Evaluative certainty → Math ability beliefs		Estimated indirect effect		95% BC CI
	B	*SE*	*B*	*SE*	*B*	*SE*	
Women							
Poor perceived performance	0.865^{*}	.375	-0.308^{***}	.074	**−0.277**	.132	[−0.6051, −0.0717]
Strong perceived performance	−0.477	.516	0.282^{**}	.104	−0.133	.142	[−0.4722, 0.1055]
Men							
Poor perceived performance	0.033	.357	-0.308^{***}	.074	−0.011	.123	[−0.2600, 0.2378]
Strong perceived performance	0.626	.475	0.282^{**}	.104	0.174	.138	[−0.0405, 0.5181]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
* $p < .05$. ** $p < .01$. *** $p < .001$.

negative stereotype converges with the valence of performance-related perceptions. Women who viewed their performance as poor were more certain of this evaluation when they reported their gender after the test compared with when they did not. However, this manipulation had no effect on men who thought they performed poorly or among participants who believed they did well on the test, regardless of their gender.

It is important to note that these differences in evaluative certainty emerged without any differences in the valence of participants' ratings of performance. As previously described, responses on the index of scaled (11-point) perceived performance perceptions did not differ as a function of gender, condition, or any interaction that included these factors. Taken together, this provides further support that postperformance stereotype accessibility can primarily serve to reinforce or validate—rather than polarize or change—existing performance evaluations.

Study 2

Study 2 was designed to extend the previous findings in some key directions. In particular, this investigation used a different manipulation of stereotype accessibility. Also, along with effects on ability beliefs, two other potential consequences of stereotype validation were examined—lowered expectations for future performance and reduced interest in math-related careers. Lastly, to focus on the negative impact of stereotype validation regarding math, only female participants were recruited.

Method

Participants and design. One hundred forty-seven female undergraduates at a large Midwestern U.S. University participated and received partial course credit. Similar to Study 1, approximately 80% of participants (121 of 147) were White. The study represented a 2 (performance evaluation: poor, strong) × 2 (gender stereotype information: absent, present) between-participants design.

Procedure and materials. The procedure was identical to Study 1, with the following exceptions. The accessibility of gender stereotypes after the test was varied using a manipulation that was modified from past research (see Spencer et al., 1999). In particular, participants in the gender stereotype-present condition read the following:

> As you may know, some past research indicates that men and women perform differently on tests of mathematical ability. However, we are still researching to determine which mental processes affect this gender difference in performance. The research you are participating in is aimed at a better understanding of this.

Participants in the gender stereotype-absent condition received no information. The dependent measures also differed somewhat from Study 1. First, the two 11-point perceived performance measures were given immediately before, rather than after, the evaluative certainty items. Also, participants reported their expectations for future performance and interest in math-related careers.

Dependent measures. Identical scaled measures of evaluative certainty (α = .95), perceived performance (α = .93), and ability beliefs (α = .93) from Study 1 were used. In addition, participants responded to 11-point scales corresponding to future performance (α = .96) and career interests (α = .93). Expectations for future performance were assessed on the following items: "Imagine taking a similar math test in the future. I predict that my performance would be": (1 = *very poor* to 11 = *very strong*) and ". . . How well do you think you would perform?" (1 = *would perform very poorly* to 11 = *would perform very well*). The career interest measures were "I am interested in pursuing a career that requires strong math skills" (1 = *strongly disagree* to 11 = *strongly agree*); "I am NOT interested in pursuing a career that requires strong math skills" (1 = *strongly disagree* to 11 = *strongly agree*; reverse-scored); "I am interested in pursuing a career that relies more on writing and verbal skills rather than math skills" (1 = *strongly disagree* to 11 = *strongly agree*; reverse-scored); and "I am interested in pursuing a career that relies more on math skills rather than writing and verbal skills" (1 = *strongly disagree* to 11 = *strongly agree*).

Results

Perceived math performance. As in Study 1, approximately two thirds of participants rated their performance as "poor" and one third reported "strong" performance on the dichotomous performance evaluation item. A 2 (performance evaluation: poor, strong) × 2 (gender stereotype information: absent, present) between-participants ANOVA was conducted on the index of responses to the two 11-point perceived performance scales. A main effect of the dichotomous performance evaluation (M_{poor} =

4.32 [$SD = 1.53$] vs. $M_{strong} = 7.42$ [$SD = 1.22$]) was found, $F(1, 143) = 150.45$, $p < .001$, $r = .72$. No other effects emerged ($ps > .33$).

Actual math performance. A two-way ANOVA on the number of correct answers showed a main effect of the dichotomous perceived performance classification ($M_{poor} = 2.89$ [$SD = 2.09$] vs. $M_{strong} = 5.37$ [$SD = 2.69$]), $F(1, 143) = 36.82$, $p < .001$, $r = .45$. No other effects were found ($ps > .18$).

Evaluative certainty. A two-way ANCOVA was performed that controlled for actual performance on the math test ($M = 3.71$, $SD = 2.58$). The hypothesized Performance Evaluation × Gender Stereotype interaction was found, $F(1, 142) = 6.28$, $p = .013$, $r = .21$ (see Figure 2).[3] Participants who viewed their performance to be poor were more certain of this evaluation when the gender stereotype information was present (adjusted $M = 7.99$, $SE = .27$) rather than absent (adjusted $M = 6.57$, $SE = .28$), $F(1, 142) = 13.33$, $p < .001$, $r = .29$. Conversely, certainty did not differ as a function of the stereotype manipulation among participants who evaluated their performance as strong (adjusted $M_{present} = 6.92$ [$SE = .40$] vs. adjusted $M_{absent} = 7.26$ [$SE = .40$]), $F < 1$. In addition to the two-way interaction, the main effect of actual math performance tended to influence certainty, $F(1, 142) = 3.68$, $p = .057$ (all remaining $ps > .14$).

Direct and indirect effects on math-related beliefs and interests.

Math ability beliefs. An ANCOVA conducted on the ability beliefs composite showed effects that were consistent with those found on evaluative certainty. For participants who perceived poor performance, ability beliefs were significantly lower when the gender stereotype was activated after the test (adjusted $M = 5.17$, $SE = .30$) compared with when it was not (adjusted $M = 6.11$, $SE = .31$), $F(1, 142) = 5.05$, $p = .026$, $r = .19$. However, the condition had no effect for those who viewed their performance as strong (adjusted $M_{present} = 7.79$ [$SE = .44$] vs. adjusted $M_{absent} = 7.69$ [$SE = .44$], $F < 1$), Performance Evaluation × Gender Stereotype, $F(1, 142) = 2.02$, $p = .158$. A main effect of participants' performance evaluation was significant, $F(1, 142) = 26.20$, $p < .001$, $r = .39$, and the covariate of actual math performance carried a marginal influence, $F(1, 142) = 3.77$, $p = .054$. The main effect of the stereotype manipulation did not approach significance ($p = .251$).

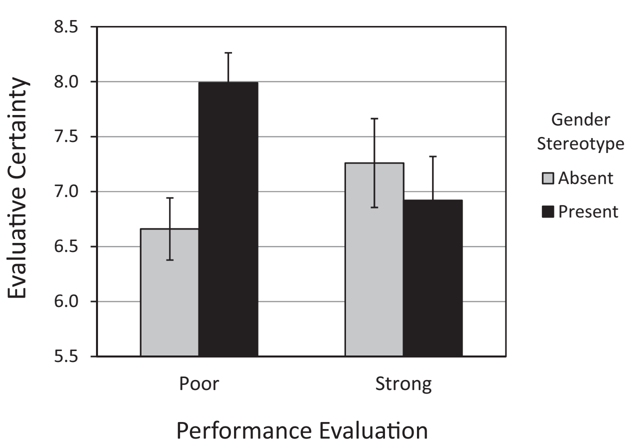

Figure 2. Adjusted mean evaluative certainty in Study 2 as a function of participants' performance evaluation and the gender stereotype manipulation (controlling for actual performance on the math test).

As in Study 1, PROCESS (Hayes, 2014) was used to test the indirect effect of evaluative certainty on ability beliefs, career interests, and future performance estimates (while controlling for actual math test performance). Moreover, the utilized model (PROCESS Model 58) allowed for the influence of the stereotype manipulation on certainty and each math-relevant outcome to depend on participants' performance evaluation (poor or strong perceived performance). The results are displayed in Table 2. With regard to math ability beliefs, findings were consistent with predictions and the results of Study 1. For women who rated their test performance as poor, the stereotype manipulation had a significant effect on their evaluative certainty and these differences were a significant predictor of decreased beliefs about one's math ability. Moreover, this indirect effect of the stereotype manipulation was found to be significant. Conversely, for women who believed they had performed well on the test, evaluative certainty predicted higher ability beliefs. However, the stereotype manipulation did not have an indirect effect.

Math career interests. The pattern found on the index of career interests tended to converge with those found on certainty and ability beliefs. For participants who evaluated their performance as poor, an ANCOVA showed that career interests tended to be lower when the gender stereotype was present after the test (adjusted $M = 3.45$, $SE = .40$) compared with absent (adjusted $M = 4.28$, $SE = .41$). However, this difference was not significant, $F(1, 142) = 2.08$, $p = .151$. For participants who perceived strong performance, no influence of the manipulation was observed (adjusted $M_{present} = 5.96$ [$SE = .59$] vs. adjusted $M_{absent} = 5.71$ [$SE = .59$], $F < 1$), Performance Evaluation × Gender Stereotype, $F(1, 142) = 1.23$, $p = .269$. A main effect of the performance evaluation was also found, $F(1, 142) = 12.94$, $p < .001$, $r = .29$, however all remaining effects were nonsignificant ($ps > .54$).

In addition to these influences, the stereotype manipulation was found to carry an indirect effect on career interests among women who believed they had performed poorly (see Table 2). For these participants, enhanced certainty as a function of the stereotype manipulation was a marginally significant predictor of lowered interest in math-related careers. Overall, however, this indirect effect was statistically reliable. On the other hand, for women who viewed their performance as strong, higher evaluative certainty predicted greater career interest—but the stereotype manipulation did not have an indirect effect.

Future performance expectations. An ANCOVA conducted on future performance estimates showed a significant main effect of participants' performance evaluation (adjusted $M_{poor} = 5.91$ [$SE = .20$] vs. adjusted $M_{strong} = 8.22$ [$SE = .29$]), $F(1, 142) = 30.45$, $p < .001$, $r = .42$. Also, a main effect of the actual performance covariate was found, $F(1, 142) = 5.65$, $p = .019$, $r = .20$. However, no additional effects emerged ($ps > .76$).

Bootstrapping analyses showed patterns of results that were consistent with those found on math ability beliefs and career

[3] The predicted Performance Evaluation × Gender Stereotype interaction also emerged from an ANOVA when actual performance was not controlled, $F(1, 143) = 5.23$, $p = .024$, $r = .19$.

Table 2
Results of Bootstrapping Moderated Mediation Analyses in Study 2

Variable	Gender stereotype manipulation → Evaluative certainty		Evaluative certainty → Math-relevant outcomes		Estimated indirect effect		
	B	*SE*	*B*	*SE*	*B*	*SE*	95% BC CI
Math ability beliefs							
Poor perceived performance	1.334***	.384	−0.529***	.096	**−0.695**	.239	[−1.2672, −0.3002]
Strong perceived performance	−0.345	.548	0.376*	.149	−0.129	.201	[−0.5270, 0.2718]
Math career interests							
Poor perceived performance	1.334***	.384	−0.254†	.138	**−0.314**	.206	[−0.8479, −0.0075]
Strong perceived performance	−0.345	.548	0.478*	.214	−0.163	.282	[−0.9132, 0.2562]
Future performance expectations							
Poor perceived performance	1.334***	.384	−0.489***	.083	**−0.693**	.233	[−1.2408, −0.3034]
Strong perceived performance	−0.345	.548	0.400**	.130	−0.141	.227	[−0.6524, 0.2384]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

interests (see Table 2). In particular, among women who rated their test performance as poor, the manipulation was found to have a significant indirect effect on expectations through differences in evaluative certainty. However, no evidence of mediation was found for women who believed they had performed well on the math test.

Discussion

The results of Study 2 offered further evidence that postperformance activation of stereotypes can validate perceptions. Consistent with the results of Study 1, women who evaluated their performance as poor were more certain of this view when the gender stereotype about math was made salient compared with when it was not. In turn, this validation triggered by stereotypes predicted decreased math-related ability beliefs, career interests, and expectations about their performance in the future. Conversely, no effects of the stereotype manipulation were found among women who evaluated their test performance as strong.

Study 3

Building upon the previous studies, the aim of Study 3 was to examine stereotype validation in the context of race and intellectual performance. Prior research has demonstrated that stereotypic beliefs of Asians tend to be associated with high competence, whereas stereotypes of Blacks correspond to relatively low competence (Fiske, Cuddy, Glick, & Xu, 2002). In particular, these stereotypes have been shown to emerge in academic performance contexts and are held by students of different races—including Black student respondents in previous research (Kao, 2000). For example, in contexts that emphasize science, technology, engineering, and math (i.e., STEM), Barden, Maddux, Petty, and Brewer (2004) found more favorable performance stereotypes for Asians compared with Blacks on both automatic and controlled measures. However, this racial difference did not emerge in non-STEM contexts. Thus, differential academic performance stereotypes for Asians versus Blacks in STEM disciplines provide a clear opportunity to establish the generality of stereotype validation phenomena.

Method

Participants and design. One hundred fifty participants at a large Historically Black University in the Eastern United States participated and received partial course credit. A total of 30 participants were dropped because they did not self-identify as African American. Furthermore, a funnel debriefing assessed whether participants believed their scores were being compared with other test takers and to indicate the comparison test takers' race. Participants who indicated disbelief in the former or could not accurately report the latter were removed from the sample. As a result of the procedures, the final sample included a total of 99 participants (73 women and 26 men). The study represented a 2 (performance evaluation: poor, strong) × 2 (racial stereotype information: absent, present) between-participants design.

Procedure and materials. The procedure paralleled Study 2, with the following exceptions. Between one and five participants arrived for each session and were greeted by a Black experimenter. Participants completed the study in one of five visually isolated cubicles within a larger room. Following the same problem-solving instructions used in Studies 1 and 2, participants completed a set of 12 challenging math questions taken from GRE practice tests. With the exception of one item, these questions were different from those used in Studies 1 and 2. Some early pretesting at the same institution as the current study suggested that the test materials used in Studies 1 and 2 would be perceived as exceptionally difficult—with a very small number of participants likely to evaluate their test performance as strong. Therefore, in an effort to achieve more balance in performance evaluations, we used a set of test questions that was somewhat less difficult (approximately 70% accuracy from past examinees, see Educational Testing Service, 1998) than the test used in the previous studies.

After the test and the dichotomous measure of perceived performance, participants received information that was designed to manipulate the accessibility of race stereotypes. In the stereotype-present condition, participants were told the following (adapted from a manipulation used by Aronson et al., 1999):

> As you may know, at many schools Asian students outnumber Black students in math majors and majors with math as a prerequisite, and

> there seems to be a growing gap in academic performance between these groups. A good deal of research indicates that Asians consistently score higher than Blacks on standardized tests of math ability.

Following this information, participants viewed a screen that listed the names of several supposed participants and were told that their performance on the math questions would be compared with that of Asian participants. The fictitious names that appeared on the screen were created to represent a mix of presumably Black and Asian students.

Conversely, participants in the stereotype-absent condition received the following information: "As you may know, some past research indicates that certain groups of students perform differently on tests of mathematical ability. However, we are still researching to determine which mental processes affect these differences in performance."

Then, these participants were given a list of names of presumably Black students only and were told their performance would be compared with other Black participants in the study.

Dependent measures. The dependent measures were identical to those used in Study 2. Furthermore, separate indices of evaluative certainty ($\alpha = .86$), perceived performance ($\alpha = .94$), ability beliefs ($\alpha = .90$), future performance ($\alpha = .99$), and career interests ($\alpha = .86$) were created by averaging responses on the pertinent items. The one exception was that the 11-point scaled measures of perceived performance were completed directly after, rather than before the measures of evaluative certainty (as in Study 1).

Results

Perceived math performance. Approximately two thirds of the participants rated their performance as "poor" and one third reported "strong" performance on the dichotomous performance evaluation item—as in Studies 1 and 2. On the 11-point index of perceived performance, a 2 (performance evaluation: poor, strong) × 2 (racial stereotype information: absent, present) between-participants ANOVA revealed a main effect of the dichotomous performance evaluation ($M_{poor} = 4.77$ [$SD = 2.03$] vs. $M_{strong} = 8.14$ [$SD = 1.92$]), $F(1, 95) = 121.34$, $p < .001$, $r = .75$. No additional effects emerged ($ps > .19$).

Actual math performance. A two-way ANOVA on the number of correct answers showed a main effect of the dichotomous perceived performance classification ($M_{poor} = 4.77$ [$SD = 2.03$] vs. $M_{strong} = 8.14$ [$SD = 1.92$]), $F(1, 95) = 57.47$, $p < .001$, $r = .61$. No other effects were found ($ps > .20$).

Evaluative certainty. To control for actual performance on the math test, the remainder of the analyses were conducted using a 2 (performance evaluation: poor, strong) × 2 (racial stereotype information: absent, present) between-participants ANCOVA. The remainder of the results were evaluated at the mean for actual performance, ($M = 5.76$, $SD = 2.52$). The predicted Performance Evaluation × Racial Stereotype interaction was found, $F(1, 94) = 4.57$, $p = .035$, $r = .22$ (see Figure 3).[4] For participants who rated their performance as poor, evaluative certainty was higher when the racial stereotype information was present, (adjusted $M = 8.01$, $SE = .38$) rather than absent (adjusted $M = 5.97$, $SE = .38$), $F(1, 94) = 13.30$, $p < .001$, $r = .35$. However, for participants who viewed their performance as strong, certainty did not differ as a function of the stereotype manipulation (adjusted $M_{present} = 8.42$ [$SE = .65$] vs. adjusted $M_{absent} = 8.45$ [$SE = .61$]), $F < 1$. Main effects of participants' performance evaluation, $F(1, 94) = 5.68$, $p = .019$, and the stereotype manipulation, $F(1, 94) = 4.43$, $p = .038$, were also found, but there was no effect of the covariate of actual math performance on certainty, $F < 1$. The results of a three-way ANCOVA that included participant gender as a factor found no significant main effect of or interactions with gender ($ps > .22$).

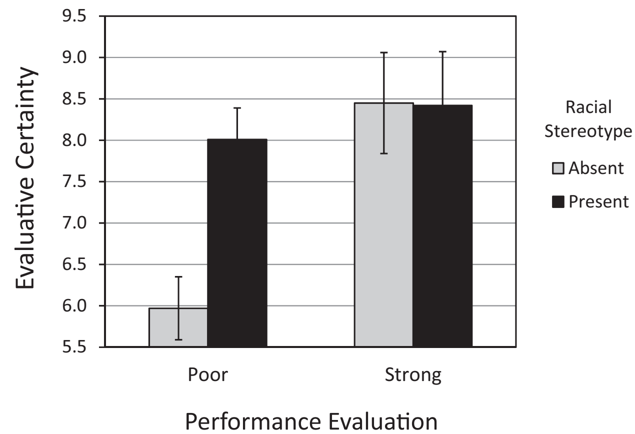

Figure 3. Adjusted mean evaluative certainty in Study 3 as a function of participants' performance evaluation and the racial stereotype manipulation (controlling for actual performance on the math test).

Direct and indirect effects on math-related beliefs and interests.

Math ability beliefs. The results of an ANCOVA revealed a pattern of effects that was consistent with the findings on evaluative certainty. Among participants who rated their performance as poor, ability beliefs were significantly lower when the racial stereotype information was present (adjusted $M = 4.74$, $SE = .32$) compared with absent (adjusted $M = 5.84$, $SE = .33$), $F(1, 94) = 6.32$, $p = .013$, $r = .25$. Conversely, this manipulation did not influence the ability beliefs of participants who perceived their performance as strong ($M_{present} = 6.57$ [$SE = .55$] vs. $M_{absent} = 7.01$ [$SE = .52$], $F < 1$). However, no Performance Evaluation × Racial Stereotype interaction was observed ($F < 1$). The main effect of participants' performance evaluation was statistically significant, $F(1, 94) = 8.39$, $p = .005$, $r = .29$, whereas the main effect of the stereotype manipulation was marginally significant, $F(1, 94) = 3.54$, $p = .063$, $r = .19$. The covariate of actual performance also predicted ability beliefs, $F(1, 94) = 15.97$, $p < .001$, $r = .38$.

Potential indirect effects on ability beliefs, career interests, and expectations of future performance were tested with bootstrapping procedures that were identical to those used in Study 2. The results are displayed in Table 3. As previously described, the racial stereotype manipulation had a significant effect on the evaluative certainty of participants who viewed their performance as poor, but not among those who thought they did well on the test. For the

[4] A two-way ANOVA that did not control actual test performance showed the predicted Performance Evaluation × Racial Stereotype interaction, $F(1, 95) = 4.97$, $p = .028$, $r = .22$.

Table 3
Results of Bootstrapping Moderated Mediation Analyses in Study 3

Variable	Racial stereotype manipulation → Evaluative certainty		Evaluative certainty → Math-relevant outcomes		Estimated indirect effect		
	B	*SE*	*B*	*SE*	*B*	*SE*	95% BC CI
Math ability beliefs							
Poor perceived performance	2.040^{***}	.525	-0.318^{***}	.082	**−0.562**	.244	[−1.1709, −0.1819]
Strong perceived performance	−0.025	.806	0.715^{***}	.209	−0.018	.453	[−0.8865, 0.9180]
Math career interests							
Poor perceived performance	2.040^{***}	.525	−0.102	.124	−0.173	.258	[−0.8400, 0.2247]
Strong perceived performance	−0.025	.806	0.718^{*}	.317	−0.018	.507	[−1.080, 1.0299]
Future performance expectations							
Poor perceived performance	2.040^{***}	.525	-0.286^{**}	.094	**−0.601**	.308	[−1.4318, −0.1471]
Strong perceived performance	−0.025	.806	0.335	.241	−0.008	.229	[−0.4116, 0.5165]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
* $p < .05$. ** $p < .01$. *** $p < .001$.

former group, increased certainty elicited by the stereotype manipulation was found to predict lower math ability beliefs and this indirect effect was significant. For participants who rated their test performance as strong, greater certainty in this evaluation predicted higher ability beliefs. However, no indirect effect of the racial stereotype manipulation was found.

Math career interests. For participants who rated their performance as poor, mean interest in math-related careers tended to be lower when the racial stereotype information was present (adjusted $M = 3.39$, $SE = .46$) rather than absent (adjusted $M = 4.10$, $SE = .47$). However, this difference did not approach statistical significance, $F(1, 94) = 1.34$, $p = .25$, $r = .12$. Among participants who evaluated their performance as strong, the stereotype manipulation had no influence on interests (adjusted $M_{\text{present}} = 5.34$ [$SE = .78$] vs. adjusted $M_{\text{absent}} = 4.72$ [$SE = .74$], $F < 1$), and no Performance Evaluation × Racial Stereotype interaction was observed, $F(1, 94) = 1.30$, $p = .26$, $r = .12$. Finally, a marginal effect of participants' dichotomous performance evaluation emerged, $F(1, 94) = 3.04$, $p = .08$, $r = .18$, and no main effect of the stereotype manipulation was found ($F < 1$). Beyond these direct influences, higher levels of evaluative certainty predicted greater interest in math-related careers among participants who viewed their test performance as strong (see Table 3). However, the bootstrapping analyses did not reveal any significant indirect effects of the racial stereotype manipulation on this dependent measure.

Future performance expectations. A two-way ANCOVA showed a significant main effect of participants' performance evaluation (adjusted $M_{\text{poor}} = 6.30$ [$SE = .25$] vs. adjusted $M_{\text{strong}} = 8.11$ [$SE = .46$]), $F(1, 94) = 10.02$, $p = .002$, $r = .31$. Also, the covariate of actual performance influenced future performance expectations, $F(1, 94) = 16.18$, $p < .001$, $r = .38$. No additional effects emerged from this analysis ($Fs < 1$).

Tests of the potential indirect effect of the racial stereotype manipulation revealed patterns that were consistent with the primary hypotheses. As shown in Table 3, enhanced certainty triggered by the manipulation was associated with lowered expectations regarding future performance for participants who thought they did poorly on the math test. Moreover, this indirect effect was found to be significant. No effects emerged among participants who evaluated their math test performance as strong.

Discussion

The results of Study 3 established the generality of stereotype validation to racial stereotypes. As in the previous studies, post-performance activation of stereotypes validated perceived level of performance, specifically when the two were congruent. In Study 3, African American participants who perceived their performance as poor were more certain of this view when a racial stereotype about math ability was made salient compared with a control condition. Furthermore, in parallel to the Study 2 findings, this enhanced certainty evoked by the stereotype manipulation predicted diminished beliefs about one's math abilities and lowered expectations for future performance. A very different pattern emerged for African American participants who believed they had performed well on the math test. Consistent with hypotheses about a lack of convergence between perceived performance and a stereotype, the stereotype manipulation carried no influence on evaluative certainty or the assessed math-relevant outcomes.

Study 4

A large body of research has shown that strong evaluations persist longer over time, are more resistant to change, and more effectively guide behavior relative to evaluations that are weak (for a review, see Petty & Krosnick, 1995). Although many factors have been shown to contribute to evaluative strength, perceptions of certainty are one of the most widely researched properties. As previously discussed, a great deal of this inquiry has directly dealt with the link between certainty and future behavior. For instance, across a number of studies, findings have shown that evaluations held with greater certainty are more likely to guide future behavior (e.g., Barden & Petty, 2008; Berger & Mitchell, 1989; Bizer, Tormala, Rucker, & Petty, 2006; Fazio & Zanna, 1978; Tormala & Petty, 2002).

The findings of Studies 1–3 are consistent with this established certainty–behavior association—wherein greater certainty in poor performance was found to predict lower math-related beliefs among both female and African American participants. With this in mind, it is plausible that stereotype validation in these contexts may also carry critical implications for the types of math-relevant information a person might choose to seek out. In Study 4, after a math test and a manipulation of stereotype accessibility, female participants had the choice to read descriptions of courses that were supposedly being considered for the course catalog at their university. This set of courses featured several that were math oriented, including titles corresponding to advanced math and related career pursuits. Consistent with decreases in beliefs and interests found in Studies 1–3, we predicted that stereotype validation would be associated with a decreased likelihood for female participants to choose to read course descriptions associated with high-level math skills.

Method

Participants and design. Seventy-eight female undergraduates at a large Midwestern U.S. University participated in exchange for partial course credit. As in Studies 1 and 2, over 80% of the participants were White (66 of 78). Participants were randomly assigned to a condition where the gender stereotype about math either was (gender stereotype present) or was not (gender stereotype absent) made salient after completing a set of math problems.

Procedure, materials, and measures. The procedure was similar to those used in Studies 1–3, with a few key changes. The previous studies offer convergent support that validation can emerge primarily when stigmatized group members believe they have performed poorly rather than well on a math test. With this in mind, in the current study we attempted to further limit the likelihood of participants having positive views of their performance. All participants completed a test that was even more difficult than those used in the previous studies. In a prior pretest, this new set of 12 GRE-math questions elicited negative performance evaluations from a very large majority of participants. [5]

After these math problems and the dichotomous measure of performance evaluation used in the previous studies, the salience of gender stereotypes was varied using a manipulation adapted from previous research (Rydell et al., 2009; see also Aronson et al., 1999). Participants in the gender-stereotype-present condition read the following information:

> As you may know, at many schools male students outnumber female students in math majors and majors with math as a prerequisite, and there seems to be a growing gap in academic performance between these groups. A good deal of research indicates that males consistently score higher than females on standardized tests of math ability. But thus far, there is not a good explanation for this. The research you are participating in is aimed at a better understanding of these differences.

Participants in the gender-stereotype-absent condition received no information. Following this manipulation, participants completed the four performance-specific evaluative certainty measures used in the previous studies (e.g., "How certain are you that your performance on the test was poor?" [1 = *not at all certain* to 11 = *very certain*]).

Subsequently, participants completed a selective exposure task that was designed to assess their interest in math-related academic and career pursuits (procedures modified from those used by Sawicki et al., 2013). Participants were provided with a list of 10 titles for new courses that were supposedly under consideration for their university course catalog. They were given a maximum of 2 min to click on any of the titles and read a corresponding brief course description that would appear after each selection. Of these fictitious courses, five were math oriented. Specifically, two courses corresponded to advanced math (e.g., "Advanced Methods in Applied Mathematics"), another focused on intermediate skills (i.e., "Intermediate Trigonometry"), one title targeted rudimentary math (i.e., "Improving Basic Math Skills"), and one course was dedicated to career pursuits (i.e., "Survey of Math-Related Careers"). The five remaining courses presumably elicited an expectation of less math-related content. In particular, three courses were designed to target verbal and writing abilities (e.g., "Rhetoric and Communication"), one class was geared toward study skills (i.e., "Improving General Study Skills and Habits"), and one title focused on human relations careers (i.e., "Survey of Careers in Human Relations"). The courses that participants chose to learn more about were recorded by the computer. Lastly, at any point during the time allotted for this task, participants could click on an icon labeled "I don't want to read any more" to advance toward the remainder of the study.

Results

Perceived math performance. Consistent with the pretest findings (see Footnote 5), 69 out of 78 participants (approximately 90%) rated their performance as "poor" on the dichotomous performance evaluation item. Given the very small number of participants who rated their performance as "strong," this dichotomous performance evaluation was not used as a between-participants factor unlike the previous studies. Furthermore, data from the 9 participants who rated their performance as "strong" were excluded from all subsequent analyses.

Actual math performance. A one-way ANOVA on the number of correct answers showed no difference as a function of the stereotype manipulation ($M_{\text{present}} = 2.57$ [$SD = 1.88$] vs. $M_{\text{absent}} = 2.29$ [$SD = 1.59$]), $F < 1$.

Evaluative certainty. Responses to the four measures of evaluative certainty ($\alpha = .96$) were averaged to form a single composite. An ANCOVA that controlled for actual math performance ($M = 2.43$, $SD = 1.74$) was conducted. The predicted main effect of the stereotype manipulation emerged from this analysis, $F(1, 66) = 8.30$, $p = .005$, $r = .33$.[6] Participants reported they were more certain of their poor performance when the gender stereotype was made salient after the math test (adjusted $M = 8.23$, $SE = .35$) compared with when it was not (adjusted $M = 6.79$, $SE = .36$). The covariate of actual performance on the math test also had a significant effect, $F(1, 66) = 5.99$, $p = .017$, $r = .29$.

Direct and indirect effects on exposure to course descriptions. Several indices were created to assess participants' exposure to the various course descriptions. Furthermore, each index was calculated by dividing exposure to a particular course or set of course descriptions by the total number of descriptions that a given participant viewed ($M =$

[5] This test included seven items used in Studies 1 and 2 in addition to five new questions (approximately 50% accuracy from past examinees; see Educational Testing Service, 1998). In a pretest of these materials, approximately 80% of female participants (24 of 29) rated their performance as "poor" on the same dichotomous perceived performance measure used in Studies 1 and 2.

[6] The hypothesized main effect of the stereotype manipulation was also found when actual math performance was not used as a covariate, $F(1, 67) = 6.75$, $p = .012$, $r = .30$.

4.43, $SD = 1.97$). Our primary hypothesis was that stereotype validation should be associated with a decreased likelihood of self-exposure to high-level math courses. In addition, we believed that stereotype validation could be linked to increased selection of course titles that are consistent with stereotypes of women. In particular, we examined whether increased certainty in poor performance on math might predict greater exposure toward the rudimentary math course and courses targeting verbal/writing abilities or related career pursuits.

Primary analyses were conducted on separate indices of exposure to the math career course, the two advanced math courses, the rudimentary math course, the human relations careers course, and the three courses relevant to verbal/writing abilities.[7] First, predictions regarding exposure to the human relations careers course and the verbal/writing courses were not supported. No direct effects of the stereotype manipulation and/or indirect effects through evaluative certainty were found. However, significant effects did emerge on exposure to the math career course, advanced math courses, and rudimentary math course. The results of these analyses are presented in the subsections to follow.

Math career course. Results of a one-way ANCOVA on exposure to the math career course showed a significant main effect of the stereotype manipulation, $F(1, 66) = 6.27$, $p = .015$, $r = .29$. Consistent with predictions, mean exposure to this course was lower when the gender stereotype was present (adjusted $M = .06$, $SE = .02$) rather than absent (adjusted $M = .13$, $SE = .02$). The covariate of actual performance on the math test did not have a significant effect ($F < 1$). As previously noted, unlike Studies 1–3, our sample only included participants who believed they had performed poorly on the math test. With this in mind, the relation between evaluative certainty and course exposure could not be moderated by participants' performance evaluation of poor versus strong. Therefore, using bootstrapping tools employed in the previous studies, analyses were performed to test the potential (nonmoderated) mediational role of evaluative certainty on course exposure (PROCESS Model 4; see Hayes, 2014). The results of these analyses can be found in Table 4. For the math career course, an analysis revealed that differences in evaluative certainty elicited by the stereotype manipulation predicted decreased exposure to this course description. Furthermore, this hypothesized pattern was found to be statistically significant.

Advanced math courses. An ANCOVA on the index of exposure to the two advanced math courses revealed no direct effect of the stereotype manipulation (adjusted $M_{\text{present}} = .11$ [$SE = .03$] vs. adjusted $M_{\text{absent}} = .12$ [$SE = .03$]), $F < 1$. Actual performance on the math test (i.e., covariate) did not have a significant effect either, $F(1, 66) = 1.97$, $p = .165$. However, a bootstrapping analysis suggested that exposure to the advanced math courses may have been guided indirectly through differences in evaluative certainty. As shown in Table 4, increased certainty as a function of the manipulation was associated with less exposure to these courses and this indirect effect was significant.

Rudimentary math course. The stereotype manipulation tended to have a direct effect on exposure to the basic math course. An ANCOVA showed that participants tended to select this course more in the condition where the gender stereotype was activated (adjusted $M = .16$, $SE = .26$) compared with the control (adjusted $M = .08$, $SE = .11$), $F(1, 66) = 2.90$, $p = .093$. The influence of the actual test performance covariate, however, did not approach significance, $F(1, 66) = 1.02$, $p = .316$. A subsequent bootstrapping test was conducted to examine whether the marginal influence of the manipulation on exposure may have been driven by higher levels of certainty in poor math performance (see Table 4). Differences in evaluative certainty triggered by stereotype manipulation tended to predict greater exposure to the rudimentary course. Although this individual path was not statistically reliable, the overall pattern of mediation was found to be significant.

Discussion

Building from the previous investigations, the findings of Study 4 extended our understanding of postperformance stereotype activation in some meaningful directions. In this investigation, evidence of stereotype validation was observed using a different manipulation of stereotype salience; thereby offering additional evidence that supports the generality of this phenomenon. Furthermore, consistent with research on evaluative certainty in other domains, the data suggested that stereotype validation may carry downstream consequences for later behavior. Following the math problems and the stereotype manipulation, participants were given the opportunity to read descriptions of supposed future course offerings at their university. Analyses of participants' choices showed patterns that were consistent with hypothesized effects of stereotype validation. In particular, increased certainty triggered by stereotype activation predicted decreased exposure to courses that presumably required high math aptitude (i.e., advanced math content courses and a math career course) but greater exposure to information corresponding to basic math skills.

Study 5

Previous findings suggest that the degree of convergence between negative stereotypes and one's perceptions of performance may be a critical aspect of stereotype validation effects. In particular, in Studies 1 and 2, only female participants who believed they had performed poorly on the math task showed greater certainty when gender or the gender stereotype were made salient. By contrast, women who viewed their math performance as strong and male participants (regardless of their performance evaluation) showed no evidence of validation from the later activated stereotype. Similarly, in Study 3, only African American participants who believed they had performed poorly on the math test—and not those who thought they did well—were influenced by the stereotype manipulation.

It is plausible that other factors could also guide these fit perceptions and ultimately determine the extent of vulnerability to stereotype validation. One potential determinant could be a person's level of identification with a negatively stereotyped group. When one's self-concept is closely, rather than loosely,

[7] No specific predictions were posited for exposure to the remaining two course titles (i.e., "Improving General Study Skills and Habits," and "Intermediate Trigonometry"). Separate exploratory analyses revealed no direct or indirect effects on exposure to these courses.

Table 4
Results of Bootstrapping Mediation Analyses in Study 4

Variable	Gender stereotype manipulation → Evaluative certainty		Evaluative certainty → Course exposure		Estimated indirect effect		95% BC CI
	B	*SE*	*B*	*SE*	*B*	*SE*	
Math career course	1.440**	.499	−0.019**	.007	**−0.028**	.016	[−0.0712, −0.0064]
Advanced math courses	1.440**	.499	−0.020*	.008	**−0.029**	.017	[−0.0747, −0.0047]
Rudimentary math course	1.440**	.499	0.019	.012	**0.027**	.017	[0.0029, 0.0768]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
* $p < .05$. ** $p < .01$.

tied to their membership in a stigmatized group, stereotype accessibility may have more profound effects on evaluations, beliefs, and behaviors (see Crocker, Major, & Steele, 1998). For example, in research on stereotype threat, increased gender identification has been linked to lower math performance for women when stereotypes were salient prior to a test (e.g., Schmader, 2002). One explanation for these and similar effects on other groups (e.g., Cole, Matheson, & Anisman, 2007) centers on the idea that an accessible stereotype may be perceived as especially characteristic of the self by highly identified individuals. Thus, consistent with some prominent accounts of priming (e.g., Wheeler, DeMarree, & Petty, 2007; see also Wheeler & Petty, 2001), greater assimilation effects on subsequent behaviors are likely to emerge (cf. Schmader, Johns, & Forbes, 2008).

Based on this rationale, susceptibility to stereotype validation may also be influenced by group identification level. In the context of gender and math performance, individuals who identify more strongly as female may show more pronounced validation effects because the match or convergence between a salient negative stereotype and poor performance is facilitated. Using procedures that most closely paralleled Study 4, Study 5 examined the influence of this potential moderator on evaluative certainty and subsequent choices regarding math-related course stimuli.

Method

Participants and design. One-hundred fifty-nine female undergraduates at a large Midwestern U.S. University received partial course credit for their participation. Over 80% of the participants were White (134 of 159). Participants were randomly assigned to one of two conditions designed to manipulate gender identity salience after completing math problems.

Procedure, materials, and measures. The procedure was identical to Study 4 except for the following changes. First, participants had an unlimited amount of time to engage in the course exposure task. However, as in Study 4, participants could exit the task at any time. After the math problems, but before the measures of evaluative certainty, approximately half of the participants answered questions designed to assess the degree to which they identified with being female (adapted from Schmader, 2002; see also Luhtanen & Crocker, 1992). Each of these questions was paired with a 7-point scale from (1) *strongly disagree* to (7) *strongly agree*. The items read as follows: "Being a woman is important to my self-image"; "Being a woman is unimportant to my sense of what kind of person I am" (reverse-scored); "Being a woman is an important reflection of who I am"; and "Being a woman has very little to do with how I feel about myself" (reverse-scored). The remaining half of participants did not answer these questions prior to measurement of evaluative certainty. Rather, they completed these items near the conclusion of the study after data from all of the dependent measures had been collected. This manipulation was used to simultaneously vary the salience of the participants' gender while capturing individual differences in gender identification. Responses to the four items were reliable ($\alpha = .72$) and were summed to form a composite gender identification score ($M = 20.82$, $SD = 3.97$).

Results

Perceived math performance. Consistent with the findings of Study 4, 130 out of 159 participants (approximately 80%) rated their performance as "poor" on the dichotomous performance evaluation item. Because of the small number of participants who rated their performance as "strong," responses to this dichotomous question were not treated as a between-subjects factor. Moreover, data from the participants who rated their performance as "strong" were excluded from all subsequent analyses (as in Study 4).

Actual math performance. Centered regression analyses were conducted on the subsequent dependent measures (see Aiken & West, 1991). In each analysis, centered predictors included the timing of the gender identification questions (categorical), individual differences in responses to these items (continuous), and their interaction. A centered regression performed on the number of correct answers on the math test ($M = 3.19$, $SD = 1.97$) revealed no significant effects ($ps \geq .27$).

Evaluative certainty. An index was created by averaging the four measures of evaluative certainty ($\alpha = .95$). A centered regression was conducted that included the previously described predictors as well as a main effect of actual math performance. In support of the predictions, a main effect of the manipulation emerged such that participants were more certain they had performed poorly on the math test when the gender identity questions were given prior to rather than after the certainty measures, $b = 1.18$, $t(125) = 3.41$, $p = .001$, $r = .29$. In addition, this main effect was qualified by the hypothesized Timing Condition × Gender Identification interaction, $b = .17$, $t(125) = 2.03$, $p = .044$, $r =$

.18 (see Figure 4).[8] At relatively low levels of gender identification (−1 *SD*), the timing of the gender questions had no significant influence on evaluative certainty, $b = .47$, $t(125) = .97$, $p = .337$. However, among participants who were highly identified as female (+1 *SD*), greater certainty about poor performance was observed when the gender questions were completed before rather than after the items corresponding to evaluative certainty, $b = 1.88$, $t(125) = 3.82$, $p < .001$, $r = .32$.

Direct and indirect effects on exposure to course descriptions. Various indices of exposure to particular courses or sets of courses ($M = 5.45$, $SD = 2.16$) were created using the same procedures from Study 4. In addition, support for any direct and/or indirect effects on the independent variables only emerged on indices of exposure to the math career course and the remedial math course.

Math career course. A centered regression revealed that exposure to this course tended to differ as a function of a Timing Condition × Gender Identification interaction, $b = -.01$, $t(125) = -1.61$, $p = .110$. For participants relatively low in female identification (−1 *SD*), no effect of the timing manipulation emerged on exposure, $b = -.01$, $t(125) = -.20$, $p = .844$. Conversely, at high levels of female identification (+1 *SD*), exposure to the math career course description was lower when gender identity was made salient beforehand, $b = -.08$, $t(125) = -2.47$, $p = .015$, $r = .22$. In addition to this key simple effect, the overall main effect of the timing manipulation, $b = -.05$, $t(125) = -1.90$, $p = .060$, and actual performance on the math test, $b = .01$, $t(125) = 1.81$, $p = .073$, were marginally significant. The main effect of gender identification level was nonsignificant, $b = .00$, $t(125) = -.15$, $p = .878$.

Indirect effects of evaluative certainty on course exposure were examined using the same general approach as the previous studies. As in Study 4, only data from participants who believed they had performed poorly were analyzed. Thus, in contrast to Studies 1–3, we did not test for how participants' perceived performance evaluation of poor versus strong could serve as a moderator. Rather, the bootstrapping analyses were conducted to examine the key Study 5 prediction regarding the moderating role of gender identification. In particular, the level of gender identification was included as a moderator of the path from the timing manipulation to the proposed mediator of evaluative certainty. However, no moderation of the path between evaluative certainty and course exposure was predicted or specified in the analysis (PROCESS Model 7; see Hayes, 2014).

The results of these bootstrapping analyses are displayed in Table 5. On exposure to the math career course, results showed differential patterns that were consistent with hypotheses. In particular, among women who were relatively low in gender identification (−1 *SD*), the timing manipulation had no impact on evaluative certainty, so there was no indirect effect on exposure to the math career course. However, for women who reported high levels of identification with their gender (+1 *SD*), higher evaluative certainty evoked by the manipulation predicted decreased exposure and this indirect effect was significant.

Advanced math courses. The results of a centered regression analysis on exposure to the advanced math courses yielded no significant direct effects of the predictors (*p*s > .15). In addition, as shown in Table 5, bootstrapping analyses did not reveal a significant indirect effect at either relatively low levels (−1 *SD*) or high levels (+1 *SD*) of gender identification.

Rudimentary math course. A main effect of gender identification level emerged such that greater exposure to the basic math course was associated with higher levels of female gender identification, $b = .01$, $t(125) = 2.95$, $p = .004$, $r = .26$. However, this influence was qualified by a significant Timing Condition × Gender Identification interaction, $b = .01$, $t(125) = 2.49$, $p = .014$, $r = .22$. Among participants who were relatively low in female identification (−1 *SD*), the timing of the gender questions had no effect on exposure, $b = -.04$, $t(125) = -1.36$, $p = .177$. However, when female identification was high (+1 *SD*), greater exposure to this basic math course was observed in the condition where the gender identity questions were completed before—rather than after—the key measures of evaluative certainty and course exposure task, $b = .06$, $t(125) = 2.16$, $p = .033$, $r = .19$. No other effects were found (*p*s > .50).

Results of a bootstrapping analysis revealed the predicted conditional mediation effects on exposure to the rudimentary math course (see Table 5). For women low in gender identification (−1 *SD*), the timing manipulation did not influence evaluative certainty; thus, no indirect effect on exposure. However, among women who were highly identified (+1 *SD*), a significant mediational relationship emerged. As hypothesized, increased evaluative certainty triggered by the timing manipulation predicted greater exposure to the description of the rudimentary math course.

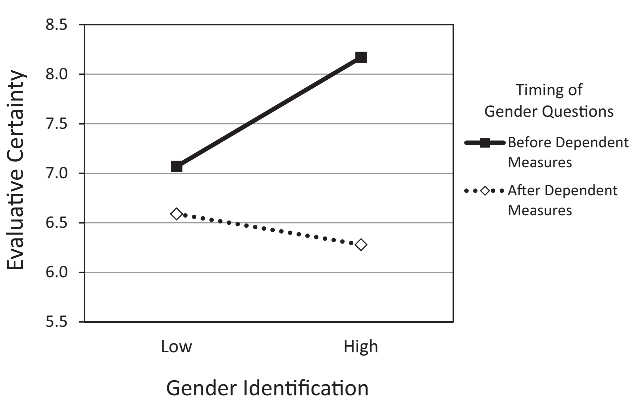

Figure 4. Study 5 predicted values for evaluative certainty as a function of the timing of the gender questions and individual differences in gender identification (controlling for actual performance on the math test). Graph is plotted at +/− 1 *SD* on the index of gender identification.

Discussion

The findings of Study 5 suggest that the extent of identification with a negatively viewed group may be a key moderator of the effects of stereotype validation. For women who believed they had performed poorly on the math test, certainty did not differ as a function of the gender salience manipulation among those with

[8] The predicted Timing Condition × Gender Identification interaction was also found via a centered regression analysis that did not control for actual performance on the math test, $b = .19$, $t(126) = 2.17$, $p = .032$, $r = .19$.

Table 5
Results of Bootstrapping Moderated Mediation Analyses in Study 5

Variable	Timing manipulation → Evaluative certainty		Evaluative certainty → Course exposure		Estimated indirect effect		95% BC CI
	B	*SE*	*B*	*SE*	*B*	*SE*	
Math career course							
Low gender identification (−1 SD)	0.470	.487	−0.012*	.006	−0.006	.008	[−0.0326, 0.0036]
High gender identification (+1 SD)	1.880***	.492	−0.012*	.006	**−0.024**	.018	[−0.0768, −0.0010]
Advanced math courses							
Low gender identification (−1 SD)	0.470	.487	−0.006	.006	−0.003	.005	[−0.0225, 0.0027]
High gender identification (+1 SD)	1.880***	.492	−0.006	.006	−0.010	.011	[−0.0360, 0.0092]
Rudimentary math course							
Low gender identification (−1 SD)	0.470	.487	0.013*	.005	0.006	.008	[−0.0069, 0.0239]
High gender identification (+1 SD)	1.880***	.492	0.013*	.005	**0.025**	.011	[0.0078, 0.0540]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
* $p < .05$. *** $p < .001$.

relatively low identification with their gender. However, for participants who reported relatively high female identification, a different pattern emerged. These women reported feeling more certain that they had performed poorly on the math test when their gender was made salient compared with the control. Furthermore, differences in certainty toward poor performance were found to hold implications for the types of math-related information participants chose to read later in the study. In particular, among women who were highly identified with their gender, greater certainty as a function of the manipulation predicted less self-exposure to the math career class and more exposure to the rudimentary math course.

One limitation of this study stems from an unexpected effect of the timing manipulation that was identified through supplementary analyses. The results of a one-way ANOVA found that gender identification differed as a function of the timing manipulation, $F(1, 128) = 8.31$, $p = .005$. Higher gender identification was reported in the condition where participants answered these questions before ($M = 21.85$, $SD = 3.96$) rather than after ($M = 19.86$, $SD = 3.90$) the dependent measures. It is unclear how this unexpected influence could have contributed to our key findings on evaluative certainty. If this main effect of the timing manipulation played a substantial role, then one might expect higher certainty when the gender identification measures were completed before rather than after the dependent measures—regardless of individual differences in gender identity. However, this pattern did not emerge. As previously discussed, the main effect of the manipulation was qualified by a significant Timing Condition × Gender Identification interaction on certainty. On the other hand, because of correlational aspects of the study, alternative explanations (e.g., those related to a disproportionate amount of highly identified participants across conditions) cannot be entirely ruled out. This issue could be addressed in future research that independently manipulates both group identification and stereotype accessibility.

Study 6

Studies 1–5 provide convergent support that stereotypes can validate perceptions when made accessible following math performance. The aim of Study 6 was to examine whether stereotypes can play a similar role in a very different performance context. Many studies have found that women are viewed as more communal, nurturing, and emotionally responsive than men (for reviews, see, e.g., Deaux & LaFrance, 1998; Eagly, Wood, & Diekman, 2000). These differences are particularly salient with regard to child care—wherein women are often believed to have greater interest in children (Prentice & Carranza, 2002), expected to serve more as primary caregivers (Hoffman & Kloska, 1995; Kellerman & Katz, 1978), and thought to possess stronger care-relevant abilities (Cejka & Eagly, 1999). Given the predominance of these beliefs, it stands to reason that making this gender stereotype accessible could validate negative perceptions related to child care performance and/or knowledge among men.

In Study 6, male and female participants answered a set of multiple-choice questions regarding infant care and development. This quiz was designed to be extremely difficult. Thus, we expected that participants would produce predominately negative performance-related reactions and view their performance as poor. Next, participants received information that was designed to manipulate the salience of the gender stereotype regarding child care. As a departure from the previous studies, no dichotomous measure was used to categorize participants' perceived performance. Relatedly, all participants also responded to the same measures of evaluative certainty—a total of two items that were similar to those used in previous studies. These changes allowed for a test of our predictions that was not contingent on the valence of performance evaluations and made the study shorter and more conducive to inexpensive online data collection.

For women, poor perceived performance in the domain of child care—unlike in math contexts— does not converge with the gender stereotype (women are good at child care). Therefore, making this stereotype accessible should not influence how certain female participants are toward perceptions of poor performance. For men, on the other hand, the gender stereotype (men are bad at child care) should converge with negative views about their performance on the quiz. Hence, we postulated that making the stereotype accessible would validate the perceptions produced by men and this increased certainty would predict lowered beliefs about their knowledge and ability in this domain.

Method

Participants and design. Three hundred sixty-six U.S. citizens (223 women and 143 men; $M_{age} = 35.01$ years, $SD = 12.40$) recruited from Amazon Mechanical Turk completed the study in exchange for $0.50. Over 70% of the participants reported they were White (261 of 366). Participants were randomly assigned to a condition in which they did or did not receive gender stereotype information after completing a quiz on infant knowledge and care. The study represented a 2 (gender: women, men) × 2 (gender stereotype information: absent, present) between-participants design.

Procedure and materials. At the start of the study, participants were informed that the researchers were interested in existing knowledge about how to care for infants. Following this information, participants completed a 10-item multiple-choice quiz on this topic that was found to yield predominantly negative evaluations of performance in a preliminary study.[9] These questions were adapted from those available on various websites and covered a range of infant-related issues corresponding to health, development, or daily care. [10] Some example questions were: "How many diapers does the average baby go through in one year?" (A = *250*, B = *2,500*, C = *5,500*, D = *7.500*, or E = *9,500*) and "What causes cradle cap?" (A = *Psoriasis*, B = *Allergies*, C = *Dry Skin*, D = *Alopecia*, or E = *No one knows for sure*). Participants provided answers to all 10 questions and were given an unlimited amount of time to do so.

Immediately after the quiz, participants rated their performance on the following 11-point scale: "Overall, how well do you think you performed on the infant-care quiz?" (1 = *performed extremely poorly* to 11 = *performed extremely well*). After this rating, participants were given information designed to manipulate the accessibility of gender stereotypes related to knowledge or care of infants/children. This induction was patterned after manipulations used in some of the previous studies. In the gender stereotype-present condition, participants received the following text: "Research suggests that women tend to perform better than men on tests of infant-care knowledge. The research you are participating in is aimed at a better understanding of this."

In contrast, participants assigned to the gender stereotype-absent condition were given this description: "Research suggests that performance on tests of infant care knowledge tends to vary as function of some personality variables. The research you are participating in is aimed at a better understanding of this."

After this manipulation, participants responded to measures of evaluative certainty, completed items targeting infant-care knowledge and ability, and answered demographic questions (gender, race, and age). Finally, participants were debriefed and thanked for their contributions.

Dependent measures. The two scaled measures of evaluative certainty were similar to items used in each of the previous studies. Each measure asked participants to "Please express how much you agree with the following statement:" (1 = *strongly disagree* to 11 = *strongly agree*) and was followed by (1) "I am certain that I performed poorly on the quiz" or (2) "I am sure that I performed poorly on the quiz." Responses to these questions ($\alpha = .94$) were averaged to form an index of evaluative certainty.

Following these measures, participants reported beliefs about their knowledge and ability with regard to caring for infants. These 11-point scales were as follows: "Please rate your own infant-care skills on the following scale" (1 = *very weak* to 11 = *very strong*); "Please rate your own infant-care knowledge on the following scale" (1 = *very low* to 11 = *very high*); "I believe that I am very knowledgeable about infant-care" (1 = *strongly disagree* to 11 = *strongly agree*); and "I believe that I could provide adequate care for an infant on my own" (1 = *strongly disagree* to 11 = *strongly agree*). Participants' ratings on each of these items ($\alpha = .93$) were averaged to form a single index of infant-care knowledge and ability beliefs.

Results

Perceived infant-care performance. A two-way ANOVA on the 11-point perceived performance scale revealed a significant main effect of gender, $F(1, 362) = 5.64$, $p = .018$, $r = .12$. Men rated their performance ($M = 4.09$, $SD = 2.38$) more negatively than women ($M = 4.69$, $SD = 2.45$). No additional effects emerged ($ps > .33$).

Actual infant-care performance. Similar to perceived performance, a two-way ANOVA on the number of correct quiz answers found a main effect of gender, $F(1, 362) = 4.91$, $p = .027$, $r = .12$. Men performed worse ($M = 2.67$, $SD = 1.30$) on the quiz than women ($M = 3.03$, $SD = 1.51$). No other effects were found ($ps > .65$).

Evaluative certainty. Analyses revealed that perceived and actual quiz performance were lower for men relative to women. Because both were assessed prior to the stereotype manipulation, analyses of the key dependent measures controlled for these differences (perceived performance $M = 4.46$, $SD = 2.44$; actual performance $M = 2.89$, $SD = 1.44$). Results of a two-way ANCOVA on evaluative certainty are displayed in Figure 5. A main effect of the stereotype manipulation emerged such that evaluative certainty was higher when the gender stereotype was made accessible after the quiz, $F(1, 360) = 7.07$, $p = .008$, $r = .14$. However this effect was qualified by the hypothesized Gender × Gender Stereotype interaction, $F(1, 360) = 4.71$, $p = .031$, $r = .11$. [11] Consistent with predictions, men reported greater certainty that they had performed poorly when the gender stereotype was present (adjusted $M = 7.41$, $SE = .18$) rather than absent (adjusted $M = 6.54$, $SE = .22$) after the quiz, $F(1, 360) = 8.97$, $p = .003$, $r = .16$. In contrast, this stereotype manipulation had no influence on the evaluative certainty of

[9] Immediately after completing the same quiz used in Study 6, nearly 75% of Amazon Mechanical Turk participants (331 out of 449) in a preliminary study rated their performance as "poor" on the same dichotomous perceived performance measure used in Studies 1–5.

[10] Questions were adapted from content available on the following websites: http://www.parents.com/parents/quiz.jsp?quizId=/templatedata/ab/quiz/data/81.xml, http://www.webmd.com/parenting/baby/rm-quiz-newborn-typical, and http://www.proprofs.com/quiz-school/quizshow.php?title=baby-shower-quiz_2&quesnum=1

[11] A two-way ANOVA that did not control for actual or perceived quiz performance revealed an interaction pattern that was directionally consistent with predictions but not significant, $F(1, 362) = 1.31$, $p = .253$. For men, mean certainty was higher when the gender stereotype was present ($M = 7.66$, $SD = 2.65$) rather than absent ($M = 7.06$, $SD = 2.58$). Conversely, for women, mean certainty was higher when the stereotype was absent ($M = 6.53$, $SD = 2.78$) rather than present ($M = 6.42$, $SD = 3.14$).

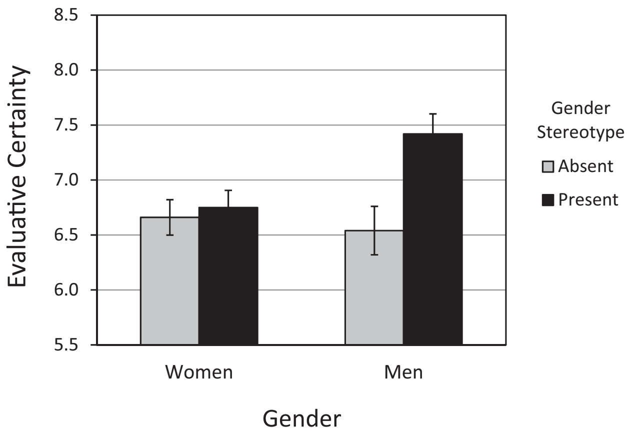

Figure 5. Adjusted mean evaluative certainty in Study 6 as a function of participant gender and the gender stereotype manipulation (controlling for actual and perceived performance on the infant-care test).

women (adjusted $M_{present}$ = 6.75 [SE = .16] vs. adjusted M_{absent} = 6.66 [SE = .16]), $F < 1$. In addition to these effects, a main effect of the perceived performance covariate was found, $F(1, 360) = 677.45$, $p < .001$, $r = .81$, but no other influences were significant (ps $> .13$).

Direct and indirect effects on infant-care beliefs. A two-way ANCOVA was performed on the composite of infant-care knowledge and ability beliefs to examine potential direct effects of the independent variables. This analysis revealed a significant main effect of gender, $F(1, 360) = 18.80$, $p < .001$, $r = .22$. Men reported lower beliefs (adjusted $M = 5.62$, $SE = .18$) than women (adjusted $M = 6.63$, $SE = .14$). Also, as with ratings of evaluative certainty, a main effect of perceived performance was found, $F(1, 360) = 203.87$, $p < .001$, $r = .60$. No other effects emerged (ps $> .13$).

The findings of Studies 1–5 suggest that stereotype validation should hold important consequences for domain-relevant beliefs and behaviors. In the current study, men were more certain that they had performed poorly on the infant-care quiz when the gender stereotype was made accessible shortly after performance. This increased evaluative certainty should, in turn, carry a negative influence on beliefs about one's knowledge and ability regarding infant/child care. On the other hand, the certainty held by women was not influenced by the stereotype manipulation. Therefore, stereotype accessibility should not carry substantial, certainty-mediated implications for the infant-care–related beliefs possessed by female participants.

As in the previous studies, a bootstrapping analysis was conducted to examine potential indirect effects of the stereotype manipulation. In particular, the analysis allowed for the influence of the stereotype manipulation on evaluative certainty to be moderated by the gender of participants. Conversely, the relationship between evaluative certainty and infant-care beliefs should not depend on participant gender (as in Study 1); thus, no such moderation was specified for this path (PROCESS Model 7; see Hayes, 2014). As displayed in Table 6, results of the bootstrapping analysis were consistent with the predicted conditional mediation effects. For women in the sample, differences in evaluative certainty were not influenced by the stereotype manipulation and did not play a mediational role on infant-care–related beliefs. For men, on the other hand, a significant indirect effect was found. Specifically, increased certainty as a function of the gender stereotype manipulation predicted lower beliefs concerning infant-care knowledge and ability.

Discussion

The findings of Study 6 indicate that stereotype validation has implications that extend beyond the domain of math performance and math-related stereotypes. Male participants were more certain that they had performed poorly on a quiz about infant care when the gender stereotype was made accessible after performance compared with a control condition. By comparison, the stereotype manipulation had no influence on the evaluative certainty reported by female participants. Critically, greater certainty elicited by the stereotype manipulation predicted diminished beliefs about one's knowledge and ability in the domain of child care. These results provide further evidence that stereotype validation occurs primarily when a stereotype converges, rather than contrasts with one's performance-related perceptions.

General Discussion

Over the past two decades, questions concerning how accessible stereotypes can influence performance have become some of the most widely studied in social psychology. Although hundreds of published studies have identified targets, various moderators, and mechanisms associated with effects of stereotype threat, no investigations to date have identified the potentially negative repercussions of stereotypes activated only after one has finished performing. Research into these possibilities seems especially relevant for standardized testing contexts. In these situations, one of the more common ways that negative stereotypes may become accessible is through the use of demographic questions that require test-takers to report their group affiliations. Based on data from a large field experiment (Stricker & Ward, 2004), some have advocated that changing the timing of demographics from before to after test completion "would be the single most cost-effective action our country could take" (Danaher & Crandall, 2008, p. 1652) to reduce group-based performance gaps in some test settings. Although this approach may hold promise for reducing stereotype threat, it could facilitate another kind of bias that could carry negative consequences.

Building from recent research on metacognition (see Briñol & Petty, 2009; Clark et al., 2009), the present studies suggest that stereotypes can validate perceptions when activated after performance. In the current research, participants completed a challenging intellectual task and rated their perceived performance. Following this evaluation, the accessibility of group stereotypes was manipulated and participants then completed measures of evaluative certainty, domain-relevant beliefs, and/or interests. Across six investigations that used different stereotype manipulations, evidence supports that accessibility of stereotypes increased the degree to which members from different negatively stereotyped groups were certain they had performed poorly on a task. In particular, Studies 1, 2, 4, and 5 found increased certainty in poor math performance for women, whereas Study 3 found parallel

Table 6
Results of Bootstrapping Moderated Mediation Analyses in Study 6

Variable	Gender stereotype manipulation → Evaluative certainty		Evaluative certainty → Infant-care beliefs		Estimated indirect effect		
	B	*SE*	*B*	*SE*	*B*	*SE*	95% BC CI
Women	0.089	.224	−0.178**	.067	−0.016	.044	[−0.1193, 0.0655]
Men	0.875**	.285	−0.178**	.067	**−0.156**	.084	[−0.3770, −0.0359]

Note. Bold indicates reliable indirect effect, where bias-corrected confidence interval (BC CI) does not include zero.
** $p < .01$.

effects among a sample of African Americans in response to math. Lastly—as an extension beyond math performance settings—men in Study 6 reported greater certainty that they performed poorly on a test of child-care knowledge when gender stereotypes were made accessible following performance. These validation effects were found to be contingent on whether one's performance evaluation (i.e., poor perceived performance) was stereotype-consistent (Studies 1, 2, and 6) and was observed to be most pronounced among participants who expressed high levels of identification with a stigmatized group (Study 5).

Consistent with research in other domains (see Petty & Krosnick, 1995; Tormala & Rucker, 2007), greater evaluative certainty—triggered by stereotype activation—was associated with key downstream consequences for domain-relevant beliefs, interests, and behaviors. Specifically, greater certainty in poor performance on a test predicted decreased beliefs about one's knowledge or abilities (Studies 1, 2, 3, and 6), lowered expectations for future performance (Studies 2 and 3), and less interest in domain-relevant career pursuits (Study 2). Finally, additional evidence suggests that stereotype validation influences the types of domain-relevant information that an individual chooses to seek out (Studies 4 and 5).

Limitations, Implications, and Future Directions

In each of the current studies, participants completed a difficult task that was designed to evoke predominantly negative reactions and evaluations of performance. These steps were taken to create circumstances that were plausibly best suited to test our primary predictions about the validating role of negative stereotypes. However, this approach was not without its limitations. First, although we attempted to limit the likelihood of stereotype threat (e.g., no mention of gender or race before a task), the performance data are consistent with these effects. In particular, men performed better than women on the math test in Study 1 and women outperformed men on the infant-care quiz in Study 6. Although the presence of these effects was not ideal, we also believe that they were largely unavoidable given one of the primary research aims—to create performance contexts that would elicit primarily negative reactions. Previous research on stereotype threat has shown that just presenting participants with a difficult task can be sufficient to produce performance differences indicative of stereotype threat (Spencer et al., 1999). It is important that in the present research, the key hypothesized effects emerged when controlling for the gender differences on performance (observed in Studies 1 and 6). Even with this in mind, however, the effects across each of the current studies should be considered in light of the possibility that some degree of stereotype threat was operating prior to and/or during task performance.

Another drawback of relying on difficult tasks in our studies was that it limited the likelihood of finding potential effects when individuals believe they performed well. Future research could address this by randomly assigning participants to conditions in which they complete either a relatively easy or difficult task prior to a manipulation of stereotype accessibility. Ultimately, such procedures could create a context to study possible validation effects among positively stereotyped individuals. If convergence between performance perceptions and a stereotype is critical, then these effects may be most likely under circumstances that differ relative to those observed in much of the current research. In particular, greater validation may occur when positive stereotypes are activated following strong perceived performance on a task.

Although the conditions that should facilitate positive validation effects were not thoroughly examined in the present studies, some data were directionally consistent with these possibilities. In Study 1—the investigation that best captured these circumstances—mean certainty among males who believed they had performed well on the math test tended to be higher when their gender was made salient compared with a control condition. However, this difference was not significant and evaluative certainty did not mediate the relation between the gender salience manipulation and math ability beliefs. This weak pattern echoes several past investigations that have examined males in math performance settings. Specifically, an accumulation of data from the stereotype threat literature suggests that activation of gender stereotypes may have some boosting influence on males (Walton & Cohen, 2003), but this effect may be very difficult to find via a single investigation and may only emerge when findings of many studies are aggregated (for a discussion, see Rydell et al., 2009).

Although evidence linked to males and math stereotypes may be difficult to obtain, previous research has identified some key ways in which positive expectancies of other groups can gain influence. For instance, several studies have shown that when the ethnic identity of Asians is activated, subsequent mathematical performance can be increased relative to a control condition (e.g., Ambady, Shih, Kim, & Pittinsky, 2001; Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999). When activated after performance, it is plausible that positive associations could also validate perceptions. For example, consistent with the convergence rationale from the current work, Asians who believed they performed well on a math task may feel more certain of this view after stereotypes of their group (e.g., Asians are good at math) are made salient.

In addition to potential positive effects, future research could also examine the specific timing conditions under which stereotype activation may facilitate validation. Previous research in persuasion suggests that variables hold the potential for validation predominantly when they are encountered after rather than before people produce primary thoughts about a stimulus. For example, Tormala et al. (2007) found that when the credibility of communicator was varied prior to receipt of a message, it affected the valence of participants' reactions toward the persuasive appeal—but not their certainty. However, when source credibility was learned only after the message, it carried a substantial influence on how certain participants were toward their message-relevant thoughts. Data from the current research are suggestive of a similar timing effect of stereotypes. As previously discussed, results on actual test performance (particularly, Studies 1 and 6) were consistent with some degree of stereotype activation prior to or during task performance—regardless of whether stereotypes were experimentally made accessible afterward or not. Yet, differences in evaluative certainty still emerged as a function of this manipulation and no group differences were found between some key conditions. For example, in Study 6, the certainty of men and women did not differ in the stereotype-absent condition. However, the gender effect on performance is consistent with the possibility that stereotypes were accessible during the infant-care quiz. Therefore, similar to past findings in persuasion contexts (e.g., Tormala et al., 2007), this suggests that potential validation might be low in cases where stereotypes become accessible prior to completion of a task.

Although the support for stereotype accessibility influencing evaluative certainty is consistent across the current studies, there is considerably less evidence for stereotypes having direct effects on the various domain-relevant outcomes. Aside from a few exceptions, the typical pattern was that the stereotype manipulation had robust indirect effects on key outcomes—through evaluative certainty—without the presence of strong direct (total) influences on these dependent variables. As discussed by Rucker et al. (2011), there are several factors that could contribute to this general type of mediation pattern. These include (but are not limited to) issues corresponding to measurement precision, having a moderate sample size, and the presence of an unmeasured suppressor variable. Of these explanations, measurement precision seems an unlikely issue considering that the statistical reliability of our measures was generally high and does not vary greatly between the mediating and dependent variables. A more tenable possibility is that some unaccounted for variable(s) could have suppressed the total direct effect between the independent and dependent variables (see MacKinnon, Krull, & Lockwood, 2000; Rucker et al., 2011). Future research could shed light on why this general ambiguity emerged in our findings by providing a more direct examination of these and other possible determinants.

Across the social psychological literature, most accumulated support for the various ways that stereotypes can gain influence has come from studies that activated stereotypes before or during a task (e.g., evaluating the attributes of another person). Although this procedure mirrors what often happens in everyday situations, there are many circumstances in which stereotypes might only become activated after the fact. Consider various electronic, media, or social networking forms of communication where important category information (e.g., race, age, and/or gender) may be obscure. Within these contexts, an individual may often produce many perceptions and form a solidified evaluation of a stimulus person before group stereotypes become salient. However, there is little empirical evidence concerning what happens when stereotype activation finally occurs. Therefore, the present studies should hold many implications and create several avenues for future inquiry that extend within and beyond the context of the reported research.

One notable direction for future endeavors should concern performance stereotypes of groups not examined in the current research. In addition to African Americans (e.g., Steele & Aronson, 1995) and women in math (e.g., Spencer et al., 1999), research has identified many different groups for which activation of negative stereotypes can inhibit later intellectual performance, including Latinos (Gonzales, Blanton, & Williams, 2002), low-SES individuals (Croizet & Claire, 1998), and U.S. Southerners (Clark, Eno, & Guadagno, 2011). It stands to reason that negative stereotypes activated after performance could impact members of these groups in ways that are similar to those found in the present research.

Future work could also examine interventions to reduce the likelihood of stereotype validation. These could include explicit postperformance efforts to reframe a previous task as being free of group differences (e.g., Spencer et al., 1999), undermining the perceived diagnosticity of a test (e.g., Steele & Aronson, 1995), or giving negatively stereotyped individuals an opportunity to affirm their own self-worth (e.g., Cohen, Garcia, Apfel, & Master, 2006; Martens et al., 2006). One additional strategy may hold especially practical implications for standardized testing. Prior research suggests that activation of positive self-relevant stereotypes can serve to counteract effects of negative associations (Rydell et al., 2009; Shih et al., 1999). For example, Rydell et al. (2009) found that stereotype threat effects were attenuated when participants' female identity and their identity as a college student were both activated via demographic questions prior to taking a math test. These researchers reasoned that the presence of the accessible positive stereotype of being a college student (i.e., college students are good at math) allowed women to rely on this identity to promote a positive self-view. It is plausible that such activation could also have a profound effect on stereotype validation. In these cases, activation of both a negative (e.g., female) and a positive (e.g., college student) identity should make perceptions of poor performance less stereotype-consistent than when only a negative identity is salient. Thus, the likelihood of high certainty in negative performance-related perceptions and the associated downstream ramifications should be reduced.

Beyond academic contexts, the reported findings should also have implications for other performance domains. Past investigations have shown that negative stereotypes can guide performance in a variety of contexts, including business negotiations (Kray, Galinsky, & Thompson, 2002), athletics (Stone, 2002), and automobile driving (Yeung & von Hippel, 2008). As with other effects of stereotype threat, performance in these studies was inhibited when negative stereotypes were accessible prior to performance. However, when activated after performance, negative stereotypes could serve to validate perceptions and, in turn, influence relevant beliefs and behaviors.

Conclusion

Across six studies, the current research demonstrated that self-relevant, negative stereotypes can have adverse effects when activated *after* intellectual performance. Specifically, for women in math (Studies 1, 2, 4, and 5), African Americans in math (Study 3), and men with regard to child-care performance (Study 6), the activation of negative group stereotypes may evoke greater certainty that one has performed poorly on a previous task. In turn, these validating effects of stereotypes were found to hold ramifications for beliefs (Studies 1, 2, 3, and 6), interests (Study 2), and behaviors (Studies 4 and 5) corresponding to the relevant performance domain. It is our hope that this research fuels future inquiry into stereotype influences on self-perceptions, beliefs, and behaviors.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12,* 385–390.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29–46. http://dx.doi.org/10.1006/jesp.1998.1371

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87,* 5–22.

Barden, J., & Petty, R. E. (2008). The mere perception of elaboration creates attitude certainty: Exploring the thoughtfulness heuristic. *Journal of Personality and Social Psychology, 95,* 489–509. http://dx.doi.org/10.1037/a0012559

Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality and Social Psychology, 71,* 637–653. http://dx.doi.org/10.1037/0022-3514.71.4.637

Berger, I. E., & Mitchell, A. A. (1989). The effect of advertising on attitude accessibility, attitude confidence, and the attitude-behavior relationship. *Journal of Consumer Research, 16,* 269–279. http://dx.doi.org/10.1086/209213

Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus online processing: Implications for attitude strength. *Journal of Experimental Social Psychology, 42,* 646–653. http://dx.doi.org/10.1016/j.jesp.2005.09.002

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 271–290). New York: Guilford Press.

Briñol, P., & Petty, R. E. (2009). Persuasion: Insights from the self-validation hypothesis. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 69–118). New York, NY: Elsevier.

Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness versus sadness as a determinant of thought confidence in persuasion: A self-validation analysis. *Journal of Personality and Social Psychology, 93,* 711–727. http://dx.doi.org/10.1037/0022-3514.93.5.711

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16,* 572–578. http://dx.doi.org/10.1111/j.0956-7976.2005.01577.x

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. http://dx.doi.org/10.1037/h0046016

Cejka, M., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to sex segregation of employment. *Personality and Social Psychology Bulletin, 25,* 413–423. http://dx.doi.org/10.1177/0146167299025004002

Clark, J. K., Eno, C. A., & Guadagno, R. E. (2011). Southern discomfort: The effects of stereotype threat on the intellectual performance of U.S. Southerners. *Self and Identity, 10,* 248–262. http://dx.doi.org/10.1080/15298861003771080

Clark, J. K., & Evans, A. T. (2014). Source credibility and persuasion: The role of message position in self-validation. *Personality and Social Psychology Bulletin, 40,* 1024–1036. http://dx.doi.org/10.1177/0146167214534733

Clark, J. K., Wegener, D. T., Briñol, P., & Petty, R. E. (2009). Discovering that the shoe fits: The self-validating role of stereotypes. *Psychological Science, 20,* 846–852. http://dx.doi.org/10.1111/j.1467-9280.2009.02375.x

Clark, J. K., Wegener, D. T., Sawicki, V., Petty, R. E., & Briñol, P. (2013). Evaluating the message or the messenger? Implications for self-validation in persuasion. *Personality and Social Psychology Bulletin, 39,* 1571–1584. http://dx.doi.org/10.1177/0146167213499238

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006, September 1). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313,* 1307–1310. http://dx.doi.org/10.1126/science.1128317

Cole, B., Matheson, K., & Anisman, H. (2007). The moderating role of ethnic identity and social support on relations between well-being and academic performance. *Journal of Applied Social Psychology, 37,* 592–615. http://dx.doi.org/10.1111/j.1559-1816.2007.00176.x

Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 504–553). New York, NY: McGraw-Hill.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype and threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24,* 588–594. http://dx.doi.org/10.1177/0146167298246003

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38,* 1639–1655. http://dx.doi.org/10.1111/j.1559-1816.2008.00362.x

Deaux, K., & LaFrance, M. (1998). Gender. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 788–827). New York, NY: McGraw-Hill.

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Mahwah, NJ: Erlbaum.

Educational Testing Service. (1998). *GRE: Practicing to take the general test: Big book*. Princeton, NJ: Author.

Fazio, R. H., & Zanna, M. P. (1978). Attitudinal qualities relating to the strength of the attitude-behavior relationship. *Journal of Experimental Social Psychology, 14,* 398–408. http://dx.doi.org/10.1016/0022-1031(78)90035-5

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28,* 659–670. http://dx.doi.org/10.1177/0146167202288010

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press.

Hayes, A. F. (2014). *PROCESS* (Version 2.12). Retrieved from http://www.afhayes.com/introduction-to-mediation-moderation-and-conditional-process-analysis.html

Hoffman, L. W., & Kloska, D. D. (1995). Parents' gender-based attitudes toward marital roles and child rearing: Development and validation of new measures. *Sex Roles, 32,* 273–295. http://dx.doi.org/10.1007/BF01544598

Jordan, A. H., & Lovett, B. J. (2007). Stereotype threat and test performance: A primer for school psychologists. *Journal of School Psychology, 45,* 45–59. http://dx.doi.org/10.1016/j.jsp.2006.09.003

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review, 2,* 137–154. http://dx.doi.org/10.1207/s15327957pspr0202_6

Kao, G. (2000). Group images and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum, 15,* 407–430. http://dx.doi.org/10.1023/A:1007572209544

Kellerman, J., & Katz, E. R. (1978). Attitudes toward the division of child-rearing responsibility. *Sex Roles, 4,* 505–512. http://dx.doi.org/10.1007/BF00287195

Kray, L. J., Galinksy, A. D., & Thompson, L. (2002). Reversing the gender gap in negotiations: An exploration of stereotype regeneration. *Organizational Behavior and Human Decision Processes, 87,* 386–410. http://dx.doi.org/10.1006/obhd.2001.2979

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18,* 302–318. http://dx.doi.org/10.1177/0146167292183006

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1,* 173–181. http://dx.doi.org/10.1023/A:1026595011371

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104. http://dx.doi.org/10.1037/1082-989X.7.1.83

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236–243. http://dx.doi.org/10.1016/j.jesp.2005.04.010

Parade, S. H., Leerkes, E. M., & Blankson, A. N. (2010). Attachment to parents, social anxiety, and close relationships of female students over the transition to college. *Journal of Youth and Adolescence, 39,* 127–137. http://dx.doi.org/10.1007/s10964-009-9396-x

Petrocelli, J. V., Tormala, Z. L., & Rucker, D. D. (2007). Unpacking attitude certainty: Attitude clarity and attitude correctness. *Journal of Personality and Social Psychology, 92,* 30–41. http://dx.doi.org/10.1037/0022-3514.92.1.30

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82,* 722–741. http://dx.doi.org/10.1037/0022-3514.82.5.722

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of metacognition in social judgment. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 254–284). New York, NY: Guilford Press.

Petty, R. E., & Krosnick, J. A. (Eds.). (1995). *Attitude strength: Antecedents and consequences*. Mahwah, NJ: Erlbaum.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42,* 185–227. http://dx.doi.org/10.1080/00273170701341316

Prentice, D. A., & Carranza, E. (2002). What women 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. http://dx.doi.org/10.1111/1471-6402.t01-1-00066

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5,* 359–371. http://dx.doi.org/10.1111/j.1751-9004.2011.00355.x

Rucker, D. D., Tormala, Z. L., Petty, R. E., & Briñol, P. (2014). Consumer conviction and commitment: An appraisal-based framework for attitude certainty. *Journal of Consumer Psychology, 24,* 119–136. http://dx.doi.org/10.1016/j.jcps.2013.07.001

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96,* 949–966. http://dx.doi.org/10.1037/a0014846

Sawicki, V., Wegener, D. T., Clark, J. K., Fabrigar, L. R., Smith, S. M., & Durso, G. R. O. (2013). Feeling conflicted and seeking information: When ambivalence enhances and diminishes selective exposure to attitude-consistent information. *Personality and Social Psychology Bulletin, 39,* 735–747. http://dx.doi.org/10.1177/0146167213481388

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38,* 194–201. http://dx.doi.org/10.1006/jesp.2001.1500

Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35,* 584–596. http://dx.doi.org/10.1177/0146167208330450

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440–452. http://dx.doi.org/10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115,* 336–356. http://dx.doi.org/10.1037/0033-295X.115.2.336

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83,* 638–664.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 80–83. http://dx.doi.org/10.1111/1467-9280.00111

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445. http://dx.doi.org/10.1037/1082-989X.7.4.422

Smith, S. M., Fabrigar, L. R., MacDougall, B. L., & Wiesenthal, N. L. (2008). The role of amount, cognitive elaboration, and structural consistency of attitude-relevant knowledge in the formation of attitude certainty. *European Journal of Social Psychology, 38,* 280–295. http://dx.doi.org/10.1002/ejsp.447

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4–28. http://dx.doi.org/10.1006/jesp.1998.1373

Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629. http://dx.doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811. http://dx.doi.org/10.1037/0022-3514.69.5.797

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). New York, NY: Academic Press. doi:10.1016/S0065-2601(02)80009-0

Stone, J. (2002). Battling doubt by avoiding practice: The effect of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin, 28,* 1667–1678. http://dx.doi.org/10.1177/014616702237648

Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology, 34,* 665–693. http://dx.doi.org/10.1111/j.1559-1816.2004.tb02564.x

Tormala, Z. L., Briñol, P., & Petty, R. E. (2007). Multiple roles for source credibility under high elaboration: It's all in the timing. *Social Cognition, 25,* 536–552. http://dx.doi.org/10.1521/soco.2007.25.4.536

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of Personality and Social Psychology, 83,* 1298–1313. http://dx.doi.org/10.1037/0022-3514.83.6.1298

Tormala, Z. L., & Rucker, D. D. (2007). Attitude certainty: A review of past findings and emerging perspectives. *Social and Personality Psychology Compass, 1,* 469–492. http://dx.doi.org/10.1111/j.1751-9004.2007.00025.x

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology, 39,* 456–467. http://dx.doi.org/10.1016/S0022-1031(03)00019-2

Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The Active-Self account. *Personality and Social Psychology Review, 11,* 234–261. http://dx.doi.org/10.1177/1088868307302223

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127,* 797–826. http://dx.doi.org/10.1037/0033-2909.127.6.797

Yeung, N. C., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis and Prevention, 40,* 667–674. http://dx.doi.org/10.1016/j.aap.2007.09.003

Received November 27, 2013
Revision received December 17, 2014
Accepted December 22, 2014 ■