The Intervening Task Method: Implications for Measuring Mediation

Jeremy P. Jamieson and Stephen G. Harkins

Pers Soc Psychol Bull 2011 37: 652 originally published online 10 March 2011
DOI: 10.1177/0146167211399776

Personality and Social Psychology Bulletin
37(5) 652–661

DOI: 10.1177/0146167211399776
http://pspb.sagepub.com

The Intervening Task Method: Implications for Measuring Mediation

Jeremy P. Jamieson[1] and Stephen G. Harkins[2]

Abstract

To study mediation, investigators sometimes examine the effect of an independent variable on an unrelated filler task that precedes the focal task. This approach assumes that the same psychological process drives performance on both tasks. The authors tested this assumption in a stereotype threat paradigm by manipulating whether or not the intervening task was described as relevant to the gender-math stereotype. When performance was relevant to the stereotype, females outperformed controls on an intervening Stroop task, but not when it was irrelevant (Experiment 1). In fact, females anticipating taking a math test under threat withdrew effort and performed more poorly on the intervening task when performance was irrelevant (Experiment 2). These findings suggest that different processes may drive performance on irrelevant and relevant intervening tasks. As a result, performance on irrelevant filler tasks may actually tell scholars little about mediating mechanisms.

Keywords

mediation, stereotype threat, intervening task, withdrawal of effort, filler task

Received March 10, 2010; revision accepted December 4, 2010

In recent years, there has been a shift in psychological research from documenting the effects of situational and internal variables on outcomes to studying the mechanisms underlying these effects. This shift to process-oriented research focus has required researchers to adapt their methods to measure underlying mechanisms. One approach taken to studying mechanisms has been to examine the effect of the independent variable on an unrelated filler task that precedes the focal task (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Forbes & Schmader, in press; Inzlicht, McKay, & Aronson, 2006; Maier, Elliot, & Lichtenfeld, 2008; Schmader & Johns, 2003). This intervening task approach assumes that once the psychological state is produced, its effects can be measured at any point during the psychological experience.

For instance, Inzlicht et al. (2006) used the intervening task paradigm to examine the effect of stereotype threat—concern about confirming, as self-characteristic, a negative stereotype about one's group (Steele & Aronson, 1995)—on performance on the Stroop color-word task. In that research, participants were led to believe that they would be taking a difficult test that was either diagnostic of intellectual ability (threat) or nondiagnostic (no threat). However, it turned out that the experimenter had run out of tests and needed to make more copies. In the meantime, the participants were asked to help out with an "unrelated pilot study," the Stroop task. Participants who anticipated performing the upcoming test under stereotype threat performed more poorly on the Stroop than nonthreatened participants. The researchers argued that the negative impact of threat on performance resulted from a depletion of self-control. That is, "stigmatized individuals use and deplete self-control to manage their devalued social identity, thus leaving them less able than their non-stigmatized counterparts to engage in self-control for other things" (p. 263). Consistent with this interpretation, in their integrated process model, Schmader, Johns, and Forbes (2008) cited the above finding as supporting a working memory account of the effect of stereotype threat on task performance. The assumption is that the same process (ego depletion/working memory deficit) that would impair the performance of threatened participants on the upcoming focal task is responsible for the participants' poor performance on the intervening Stroop task.

However, the motivation-based mere effort account (Harkins, 2006; Jamieson & Harkins, 2007, 2009; McFall, Jamieson, & Harkins, 2009) suggests that on an inhibition task like the Stroop, with the task instructions used by Inzlicht et al. (2006), the experience of stereotype threat should facilitate, not impair, performance. This prediction,

[1]Harvard University, Cambridge, MA, USA
[2]Northeastern University, Boston, MA, USA

Corresponding Author:
Jeremy P. Jamieson, Harvard University, 1206 William James Hall, 33 Kirkland Street, Cambridge, MA 02138
Email: jamieson@wjh.harvard.edu

if supported, would indicate that the mechanism at work on the intervening, filler task may not be the same as the mechanism that would drive performance on the focal task.

The mere effort account argues that stereotype threat motivates participants to perform well, which potentiates whatever response is prepotent on the given task. If the prepotent response is correct, performance is facilitated. If the prepotent response is incorrect, and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. However, if participants recognize the error and have the knowledge and time needed for correction, performance can be facilitated.

In a test of this account, Jamieson and Harkins (2007, Experiment 3) examined performance on the antisaccade task, an inhibition task that requires participants to report the orientation of a target presented on one side or the other of a visual display. Prior to target onset, a cue is presented on the opposite side of the display. Participants are explicitly instructed to not look at this cue but rather to look to the opposite side of the display where the target will appear. However, there is a reflexive-like, prepotent tendency to look toward the cue that must be inhibited (or corrected) to optimize performance. When females were told that there were gender differences on the antisaccade task, which was described as a measure of "visuospatial capacity" that was diagnostic of math ability, they generated significantly more reflexive saccades toward the cue (i.e., their prepotent response was potentiated) than control females. However, because this response was obviously incorrect and participants were given the opportunity for correction, the females subject to threat launched volitional saccades—correct saccades (eye movements to the target following successful inhibition) and corrective saccade (eye movements to the target following reflexive saccades)—more quickly than controls. Finally, after the participants' eyes arrived at the target area, the participants had to determine the target orientation and press the appropriate response key. Motivation led stereotype threat participants to try to respond as quickly as possible. That is, when the participants fixated the target, the threatened participants responded more quickly than controls.[1]

The Stroop, like the antisaccade, is an inhibition task. The prepotent response, reading the word, is incorrect, and the experience of stereotype threat should potentiate this response. However, it is also obvious on the Stroop, as on the antisaccade task, that this response is incorrect—participants know they are not supposed to read the word. Given enough time for correction, participants who are more motivated should outperform controls, just as they did on the antisaccade task. Consistent with this argument, McFall et al. (2009) found that when participants subject to evaluation were given sufficient time to inhibit their prepotent tendency to read the color word (2 s) on the Stroop, they outperformed their no evaluation counterparts, but when the response window did not afford sufficient time for correction (e.g., 1 s), the direction of the effect was reversed: No evaluation participants outperformed their evaluation counterparts.

In Inzlicht et al.'s (2006) intervening task research, participants were asked to name the colors in which a series of color words were printed on a card as quickly as possible, and the researchers measured the total time it took to complete the set. Under these conditions, participants should have had the opportunity to correct for the potentiated response because there was no response window limitation. As a result, if threatened participants were more motivated than controls, they would be expected to outperform no threat participants, but they did not. However, there is no reason for participants to be motivated to perform well on an intervening task that is unrelated to the stereotype. This analysis would suggest that the process that drives performance on an intervening task that is seen as unrelated to the psychological state of interest may be different from the process in effect when performance on the task is seen as related to the psychological state. As a result, performance on nondiagnostic intervening tasks may tell us little about the mechanism responsible for performance on focal tasks.

Experiment 1

To test this possibility, we conducted an experiment with Inzlicht et al.'s (2006) intervening task method in which the Stroop was described as a filler task but also included conditions in which females were told that performance on the Stroop was directly relevant to the stereotype that males are better than females at math. Specifically, females were told that they would be asked to solve a set of GRE problems on which either gender differences had been found (threat) or no gender differences had been found (no threat). After working on some practice problems, participants performed another task (the Stroop) that they were led to believe either was unrelated to math ability (filler task) or was a measure of processing efficiency that predicted mathematical ability on which gender differences had been found (threat).

The mere effort account predicts that when the intervening task (the Stroop) is described as relevant to the stereotype, participants will outperform participants for whom the Stroop is described as unrelated to math ability, whether the anticipated math test is diagnostic (threat) or not (no threat). In the filler condition, we should replicate Inzlicht et al.'s (2006) finding: poor intervening task performance when participants anticipate threat on the upcoming test. In addition to supporting the mere effort account of the effect of threat on performance on inhibition tasks, this pattern of results would show that the assumption that underlies the use of the intervening task method may be unwarranted. That is, the psychological process produced by the manipulation of the independent variable that would drive performance on the focal task may not be the same process that underlies performance on the intervening filler task.

On the other hand, if participants anticipating threat suffer from deficits in cognitive resources as a result of worrying about performing poorly on the upcoming diagnostic task, their performance on the intervening task should be worse than that of participants not anticipating threat whether they perform the intervening Stroop task under threat or not. That is, threatened participants must cope with concern about confirming the negative stereotype on the present (intervening) and/or the upcoming focal task. Thus, whether they perform the Stroop under threat, anticipate taking the math test under threat, or both, they would be expected to perform more poorly on the intervening Stroop task than participants who are not under threat on either task.

Method

Participants. For partial fulfillment of a course requirement, 64 Northeastern University female undergraduate students participated in this experiment. All participants were native English speakers.

Procedure. Each participant was greeted by a male experimenter, given two practice GRE problems to complete, and told that she would be taking a test made up of similar problems. Then, as in past research (e.g., Jamieson & Harkins, 2007; O'Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999), the participant was told that gender differences either did (stereotype threat) or did not (no stereotype threat) exist on this test. Participants were left to infer that men outperformed women based on the societal stereotype. At this point, as in Inzlicht et al.'s (2006) research, the experimenter discovered that he had run out of copies of the math test and asked the participant if she would complete another task (the Stroop) while copies were being made. All participants agreed to take part.

The experimenter then implemented the threat manipulation for the Stroop task. In the filler condition, the Stroop was said to measure "processing efficiency" and was unrelated to math ability. In the threat condition, females were told that the Stroop was a measure of "processing efficiency" that was highly predictive of mathematical ability.

As the experimenter was leaving to presumably make copies, he took the participant to an adjoining room and introduced her to a second experimenter who was blind to her condition. The second experimenter explained to the participant that she would be presented with lists of 100 items each and that she was to name the ink colors the items were printed in as quickly and accurately as possible. Each participant read two lists. Before the participant completed each list, she was given a short list of practice items to familiarize her with the list type on which she would be working. The incongruent list consisted of a list of color words (red, blue, green) printed in incongruent colors, whereas the control list was a series of Xs printed in the same colors. The primary dependent variable was total time it took each participant to complete each list. List order was counterbalanced.

Table 1. Anticipated Threat Manipulation Checks From Experiment 1

	Anticipated threat manipulation checks			
	Gender differences		Males vs. females	
Condition	*M*	*SD*	*M*	*SD*
Anticipated threat	8.06	2.27	3.93	1.58
Anticipated no threat	4.84	2.92	5.47	1.38

After completing the Stroop, participants completed two questionnaires, one about the Stroop and the other about the anticipated GRE test to check for the success of the manipulations. Participants were asked, "Are there gender differences?" (1 = *no gender differences*, 11 = *gender differences*) and "Do males or females perform better on this task?" (1 = *males perform better*, 6 = *males and females perform the same*, 11 = *females perform better*). We also asked participants to rate the extent to which they tried as hard as possible on the Stoop task (1 = *not at all*, 11 = *very hard*) and the extent to which they expected to try "as hard as possible on the GRE test" (1 = *not at all*, 11 = *very hard*).

As in Inzlicht et al. (2006), participants never performed the GRE, and the experimenter ended the study after participants filled out the post-Stroop questionnaire.

Results

Manipulation checks. Manipulation checks were analyzed in 2 (anticipated threat: threat vs. no threat) × 2 (Stroop threat: threat vs. filler) between-subjects ANOVAs. Manipulations were successful. Participants anticipating threat reported that gender differences existed, $F(1, 60) = 23.60$, $p < .001$, $d = 1.25$, and that males outperformed females, $F(1, 60) = 17.26$, $p < .001$, $d = 1.07$, to a greater extent than no anticipated threat participants on the upcoming GRE test. Likewise, Stroop threat participants reported that gender differences existed, $F(1, 60) = 17.63$, $p < .001$, $d = 1.08$, and that males outperformed females, $F(1, 60) = 10.54$, $p = .002$, $d = 0.84$, to a greater extent than filler participants on the intervening Stroop task (see Tables 1 and 2 for means and standard deviations).

Because the Stroop is not, on its face, a math test, participants were asked to rate the extent to which the Stroop was related to mathematical ability (1 = *not at all*, 11 = *closely related*). Stroop threat females believed the Stroop was more diagnostic of math ability ($M = 6.09$, $SD = 2.48$) than females in the filler condition ($M = 3.78$, $SD = 2.31$), $F(1, 60) = 14.77$, $p < .001$, $d = 0.99$. Anticipated threat condition did not affect participants' ratings of the diagnosticity of the intervening task.

Table 2. Stroop Threat Manipulation Checks From Experiment 1

	Stroop threat manipulation checks			
	Gender differences		Males vs. females	
Condition	*M*	*SD*	*M*	*SD*
Stroop threat	7.66	1.81	4.81	2.66
Stroop filler	5.15	2.88	6.75	2.03

Stroop performance. Researchers often use interference scores (time on incongruent list – time on control list) to analyze Stroop performance. However, if experiencing threat on the Stroop motivates participants, this effect may be present for both the control and experimental stimuli, as has been found in previous research on the effect of evaluation threat on Stroop performance (McFall et al., 2009). Thus, motivational effects can potentially be missed if interference scores are used (using the control list as a covariate creates the same problem). To avoid this problem, we analyzed the time taken to read the control and incongruent lists in a 2 (anticipated threat) × 2 (Stroop threat) × 2 (list type: incongruent vs. control) mixed-factor ANOVA. This analysis loses none of the information captured by interference scores while retaining information that would otherwise be lost.

This analysis produced a Stroop Threat × Anticipated Threat × List Type interaction, $F(1, 60) = 3.07$, $p = .08$, $d = 0.45$. To test the predictions of the mere effort and working memory accounts, we split the three-way interaction into two 2 (Stroop threat) × 2 (anticipated threat) analyses: one for incongruent trials and one for control trials. Pairwise contrasts were used decompose significant interactions (Kirk, 1995).

Analysis of the incongruent trials yielded a Stroop Threat × Anticipated Threat interaction, $F(1, 60) = 9.17$, $p = .003$, $d = 0.78$. Replicating Inzlicht et al. (2006), when participants believed the intervening Stroop task was unrelated to the stereotype, females anticipating threat on the GRE performed more poorly than those not anticipating threat, $F(1, 60) = 19.28$, $p < .001$, $d = 1.13$. However, when the intervening task was presented as diagnostic of mathematical ability, the effect of anticipated threat was eliminated. In fact, as shown in Figure 1, the performance of the Stroop threat participants taken together was better than that of either the anticipated threat/filler participants, $F(1, 60) = 42.33$, $p < .001$, $d = 1.68$, or the no anticipated threat/filler participants, $F(1, 60) = 6.48$, $p = .014$, $d = 0.66$.[2]

The two-way ANOVA performed on the control trials produced a main effect for Stroop threat, $F(1, 60) = 6.82$, $p = .011$, $d = 0.67$. Participants in the Stroop threat condition completed the control stimuli more quickly ($M = 58.53$ s, $SD = 10.26$) than filler participants ($M = 64.16$ s, $SD = 8.44$). This outcome is consistent with the mere effort account, which would suggest that the threatened participants would be more motivated than filler participants (cf. Jamieson & Harkins, 2007; McFall et al., 2009).

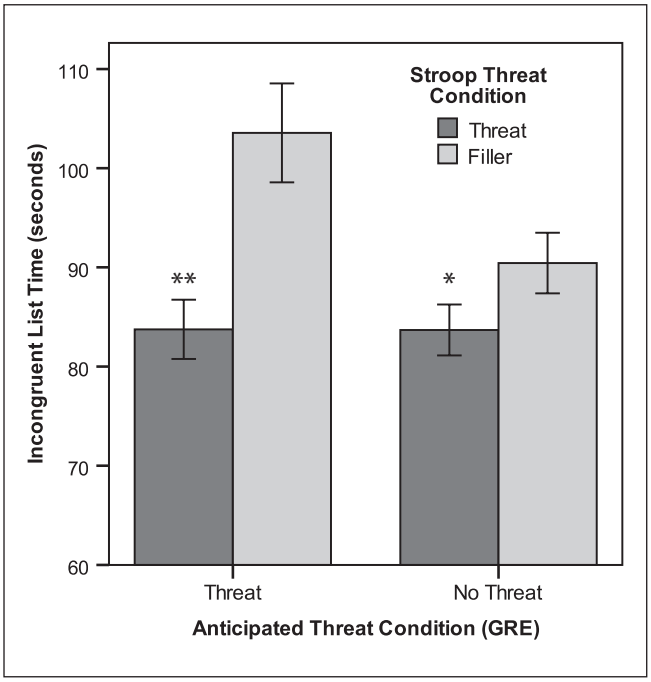

Figure 1. Time to complete the incongruent list in seconds as a function of the anticipated threat and Stroop threat conditions
Note: Error bars = +/- standard error of the mean.
*Means within the anticipated threat condition differ at $p < .05$. **Means within the anticipated threat condition differ at $p < .01$.

Ancillary measures. Self-reports of effort were analyzed in the same 2 (Stroop threat) × 2 (anticipated threat) between-subjects ANOVA that was used to analyze the manipulation check items. Participants anticipating threat on the GRE test reported putting out less effort on the intervening Stroop task ($M = 8.97$, $SD = 1.71$) than those not anticipating threat ($M = 9.84$, $SD = 1.27$), $F(1, 60) = 5.45$, $p = .023$, $d = 0.60$. Females in the anticipated threat condition also reported that they expected to put less effort into GRE performance ($M = 8.63$, $SD = 1.83$) than females in the no anticipated threat condition ($M = 9.50$, $SD = 1.14$), $F(1, 60) = 5.20$, $p = .026$, $d = 0.59$.

Discussion

Inzlicht et al. (2006) used results from the intervening task method to argue that the experience of stereotype threat depletes cognitive resources needed for self-control, just as Schmader et al. (2008) argued that threat depletes working memory, which impairs the participants' ability to inhibit response conflict. The effect of this loss should be seen in the performance of an inhibition task like the Stroop, which requires participants to stop themselves from reading the color words. Consistent with this account, in the current research we replicated Inzlicht et al.'s finding that participants who anticipated taking the GRE test under threat performed the Stroop more poorly than controls. However, this effect was obtained *only* when the Stroop was framed as an unrelated, filler task. When the Stroop was presented as a

measure of processing efficiency that predicted mathematical ability, females performed better, not worse, than controls, whether they anticipated taking the GRE under threat or not.

This pattern of performance is not consistent with the ego depletion/working memory account but instead supports predictions made by the motivation-based, mere effort account (Jamieson & Harkins, 2007, 2009). This explanation is based on the premise that stereotype threat motivates individuals to try to disprove the negative stereotypes directed at their group. This motivation, in turn, potentiates prepotent responses. However, the prepotent response is obviously incorrect on the Stroop, and the list-reading procedure used (i.e., unlimited response window) provides the more motivated threat participants with the opportunity to correct for the potentiated tendency to read the words, allowing them to complete the list more quickly than their no threat counterparts with no cost in accuracy. These findings suggest that the process that drives performance on an intervening task that is seen as irrelevant to the stereotype is different from the process in effect when performance on the intervening task is relevant to the stereotype. When the intervening task is seen as related to the stereotype, participants are motivated to perform well, and on the Stroop under the current conditions, they do so.

Although Experiment 1 shows that intervening task methods may not be ideal for studying mediating mechanisms because the process(es) that drive performance on focal tasks may differ from those responsible for unrelated, intervening task performance, this experiment was not designed to nor does it tell us why the anticipated threat participants performed more poorly on the filler task than participants who did not anticipate threat. There are at least two possible explanations for this finding. First, when the intervening task is not diagnostic of ability in the stereotyped domain, anticipated threat participants may ruminate about their upcoming performance on the diagnostic task, depleting available cognitive resources for the intervening task. That is, participants may experience a deficit in processing capacity when the intervening task is not diagnostic, even though they do not suffer this deficit when it is.

Second, anticipated threat participants may withdraw effort on intervening tasks that are not diagnostic of ability in the stereotyped domain, thereby saving resources for performance on the focal task. The self-report data from Experiment 1 could be seen as consistent with this possibility: Participants who anticipated taking the GRE under threat reported putting less effort into the Stroop task than no anticipated threat participants. However, it should be noted that this was a main effect. That is, participants who anticipated taking the GRE under threat reported putting less effort into the Stroop than participants in the no anticipated threat condition whether the Stroop was presented as a filler task or as a stereotype-relevant task. This self-report is inconsistent with the performance measure, which showed that anticipated threat/Stroop threat participants outperformed not only the anticipated threat/filler task participants but also the participants in the no anticipated threat/filler condition.

We could speculate on the source of this discrepancy, but in fact, it is not at all unusual to find a lack of correspondence between self-reports of effort and actual measures of performance. For example, the social loafing effect is highly replicable, but the average effect size for self-reported effort in these same experiments is only .07, not significantly different from zero (Karau & Williams, 1993). No reliable relationship between self-reports of effort and performance has been found in the stereotype threat literature (e.g., Brown & Pinel, 2003; Keller, 2002; Schmader & Johns, 2003; Steele & Aronson, 1995). Thus, it appears that participants are unwilling or unable to accurately report the amount of effort that they put into task performance. As a result, we cannot draw a conclusion as to whether anticipating threat depletes resources as suggested by the ego depletion/working memory account or whether it leads participants to withdraw effort.

Experiment 2

Experiment 2 sought to test between the ego depletion/working memory and withdrawal of effort accounts for the filler task performance in Experiment 1. To do so, we needed an intervening task for which the two explanations make different predictions. The Stroop task is not suitable for this purpose because it is sensitive to cognitive interference (e.g., MacLeod, 1991) as well as to differences in effort output (e.g., McFall et al., 2009). Instead, we needed a task that would be performed poorly either if participants experienced a reduction in processing capacity or if participants withdrew effort, but would not be negatively affected by both.[3]

One task that meets this requirement is the prosaccade task. On this task, participants fixate on a central crosshair prior to each trial. After a random interval, the fixation cross disappears and a cue appears on the right or left side of the screen. Participants are instructed to look at the cue. The target (an arrow in this case) then appears in the same location as the cue, and participants are told to identify the orientation of the target as quickly and accurately as possible. Of course, in this case, unlike on the antisaccade task, the prepotent response, looking at the cue, is the correct response.

Jamieson and Harkins (2007, Experiment 3) demonstrated that, on this task, effort does not affect the amount of time it takes for participants to make an eye movement to the cue/target location, but it does affect the time it takes participants to report the target's orientation once the eyes have arrived at the target site: Threatened participants reported the target orientation faster than controls. In addition, research indicates that experimentally manipulated deficits in executive processing capacity do not affect the amount of time taken to make an eye movement to the cue on the prosaccade task (Roberts, Hagar, & Heron, 1994), nor do they affect the time taken to identify the target once the eyes arrive at the target site (Jamieson & Harkins, 2007, Experiment 4). Thus,

performance on this simple visual search task is sensitive to the amount of effort exerted by participants but is unaffected by deficits in processing capacity. As a result, performance on the prosaccade task allows us to test the two explanations for the anticipated threat effect observed in Experiment 1.

In Experiment 2, we again crossed the framing of the intervening task (prosaccade threat vs. filler) and anticipated threat manipulations. This experiment has two primary goals: (a) to replicate the Stroop threat effect observed in Experiment 1 on a different task and (b) to provide an explanation as to why anticipating threat debilitates filler task performance. To test the first prediction, we simply need to examine the main effect for the intervening task frame collapsed across the anticipated threat conditions. Consistent with the findings from Experiment 1, we expected participants assigned to the prosaccade threat condition to outperform those who believed the intervening task was unrelated to the stereotype (filler task). As for the second prediction (withdrawal of effort vs. cognitive deficit), if the debilitating effect of anticipated threat in the filler condition in Experiment 1 was due to a deficit in processing capacity, this deficit should not affect performance on the prosaccade task, and we should observe only the main effect for prosaccade threat. On the other hand, if anticipating threat leads participants to withdraw effort on unrelated intervening tasks, participants performing the filler task should exhibit slower reaction times when anticipating threat than when not. This pattern of data would produce an interaction between prosaccade threat and anticipated threat.

Method

Participants. For partial fulfillment of a course requirement, 60 Northeastern University female undergraduate students participated in this experiment. All participants were native English speakers and had normal or corrected-to-normal vision.

Procedure. The procedure was identical to that used in Experiment 1 except that the participants completed the prosaccade task instead of the Stroop. All of the 120 prosaccade trials began with the presentation of a white central fixation cross on a black background for a randomly determined interval between 1,500 and 3,500 ms. The cue, a white square subtending 1° of visual angle, then appeared 11° to the left or the right for 400 ms. When the cue was extinguished, the target, an arrow also occupying 1° of visual angle, appeared in the same location for 150 ms. The target was presented in one of three orientations: pointing up, to the left, or to the right. After target offset, a mask appeared in its place and remained until participants made a response. If no response was made after 1,500 ms, the trial ended and the next trial began after a 1,750 ms intertrial interval.

Participants were instructed to indicate the orientation of the target as quickly and accurately as possible by pressing the corresponding arrow key on a keyboard. Cue side and arrow direction were randomized across trials.

Table 3. Anticipated Threat Manipulation Checks From Experiment 2

	Anticipated threat manipulation checks			
	Gender differences		Males vs. females	
Condition	*M*	*SD*	*M*	*SD*
Anticipated threat	7.37	2.20	3.73	1.41
Anticipated no threat	4.23	2.43	5.67	1.17

Table 4. Prosaccade Threat Manipulation Checks From Experiment 2

	Prosaccade threat manipulation checks			
	Gender differences		Males vs. females	
Condition	*M*	*SD*	*M*	*SD*
Prosaccade threat	5.93	2.26	4.20	1.69
Prosaccade filler	3.83	3.02	5.87	2.08

Results

All data were analyzed in 2 (anticipated threat) × 2 (prosaccade threat) between-subjects ANOVAs.

Manipulation checks. As in Experiment 1, prosaccade and anticipated threat manipulations were successful. Participants anticipating threat reported that gender differences existed, $F(1, 56) = 26.48$, $p < .001$, $d = 1.37$, and that males outperformed females, $F(1, 56) = 29.09$, $p < .001$, $d = 1.44$, to a greater extent than no anticipated threat participants on the upcoming GRE test. Likewise, prosaccade threat participants reported that gender differences existed, $F(1, 56) = 9.10$, $p = .004$, $d = 0.81$, and that males outperformed females, $F(1, 56) = 11.39$, $p = .002$, $d = 0.90$, to a greater extent than participants for whom the intervening prosaccade task was described as a filler task (see Tables 3 and 4 for means and standard deviations).

Like the Stroop, the prosaccade task is not, on its face, a math test. Thus, participants were asked to rate the extent to which it was related to mathematical ability on the same 11-point scale outlined in Experiment 1. Consistent with the success of the manipulation, females in the prosaccade threat condition believed the prosaccade task was more diagnostic of math ability ($M = 6.27$, $SD = 2.55$) than females in the filler condition ($M = 3.13$, $SD = 2.21$), $F(1, 56) = 24.91$, $p < .001$, $d = 1.33$.

Prosaccade performance. As shown in Figure 2, analysis of the time participants took to identify the orientation of the

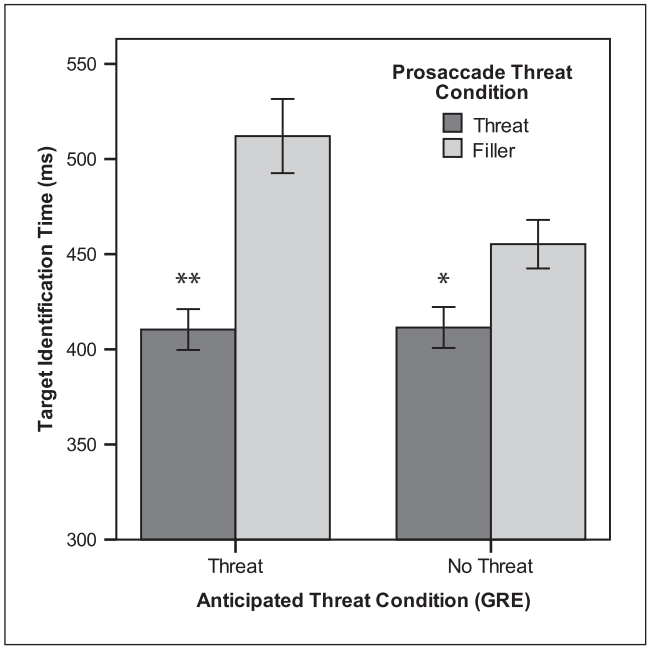

Figure 2. Time to indicate target orientation in milliseconds as a function of the anticipated threat and prosaccade threat conditions

Note: Error bars = +/- standard error of the mean.

*Means within the anticipated threat condition differ at $p < .05$. **Means within the anticipated threat condition differ at $p < .01$.

target produced a Prosaccade Threat × Anticipated Threat interaction, $F(1, 56) = 4.32$, $p = .04$, $d = 0.56$. Replicating Experiment 1, when the intervening task was presented as diagnostic of mathematical ability, the performance of the prosaccade threat participants was better than that of participants in both the filler/anticipated threat, $F(1, 56) = 32.58$, $p < .001$, $d = 1.52$, and the filler/no anticipated threat conditions, $F(1, 56) = 5.09$, $p = .028$, $d = 0.60$. Consistent with a withdrawal of effort explanation and replicating the pattern found in Experiment 1, females in the filler condition who anticipated threat performed more poorly than those not anticipating threat, $F(1, 56) = 8.39$, $p = .005$, $d = 0.77$.

Ancillary measures. Participants assigned to the anticipated threat condition reported putting out less effort on the prosaccade task ($M = 8.60$, $SD = 1.54$) than controls ($M = 9.37$, $SD = 1.38$), $F(1, 56) = 4.25$, $p = .044$, $d = 0.55$. We also observed a marginal anticipated threat effect for expected effort on the upcoming GRE test, $F(1, 56) = 3.30$, $p = .075$, $d = 0.49$. That is, participants anticipating threat predicted that they would not try as hard on the GRE ($M = 8.41$, $SD = 1.94$) as no threat participants ($M = 9.20$, $SD = 1.47$).

Discussion

Replicating the pattern of findings in Experiment 1, females informed that gender differences existed on the intervening task outperformed those assigned to the filler condition, regardless of anticipated threat condition. This finding is consistent with the growing body of evidence that argues that motivational processes play an important role in stereotype threat performance effects (e.g., Carr & Steele, in press; Hess, Hinson, & Hodges, 2009; Jamieson & Harkins, 2007, 2009; Kray, Thompson, & Galinsky, 2001; O'Brien & Crandall, 2003; Seibt & Forster, 2004).

However, we also found that when the intervening task was described as unrelated to math ability, females who anticipated taking the GRE task under threat performed the same, simple prosaccade task more poorly than females not anticipating threat. This reduction in performance on the prosaccade task would not be predicted by an ego depletion/working memory deficit explanation, because performance on this task has been shown to be unaffected by working memory capacity (Engle, 2002; Jamieson & Harkins, 2007, Experiment 4; Kane, Bleckley, Conway, & Engle, 2001; Roberts et al., 1994), unlike performance on even simple math problems (Ashcraft & Kirk, 2001). Rather, this pattern of performance is consistent with the argument that participants who are anticipating threat withdraw effort as they await their performance on the upcoming GRE test. After all, their performance on the intervening task is not relevant to the stereotype. Thus, there is no reason for them to be concerned about performing well on this task, whereas there is every reason for them to want to perform well on the upcoming GRE test.

If participants anticipating threat withdrew effort on the prosaccade task when it was framed as an unrelated filler task, then one may ask whether this withdrawal was deliberate or nonconscious. In other words, did females purposely conserve resources for the upcoming GRE test, or was the process more implicit? Females who anticipated threat on the GRE reported trying less on the prosaccade task and expected to try less on the GRE than those who did not anticipate threat. Thus, one could argue that our anticipated threat participants consciously withdrew effort. However, consistent with the findings of Experiment 1, even though participants anticipating the performance of the GRE under threat reported putting out less effort on the prosaccade task than females in the no threat condition, females performed better on that task when it was framed as a measure of math ability than when it was described as a filler task. So, as is often the case (e.g., Karau & Williams, 1993), self-reports of effort did not map onto actual performance. In sum, although the self-report data could be seen as consistent with a conscious withdrawal of effort argument, additional work (that does not rely on self-report methods) is required to elucidate this process.

General Discussion

This research shows that the processes that account for the effect of a manipulation on the focal task and an intervening

filler task can be different. In both Experiment 1 and Experiment 2, informing females that an intervening test was a measure of quantitative ability on which there were gender differences facilitated their performance, whether participants anticipated threat on an upcoming GRE test or not. However, anticipated threat impaired intervening task performance when that intervening task was presented as an unrelated filler task.

Combined with previous research (Jamieson & Harkins, 2007), the current work shows that the motivation-based mere effort account can explain how stereotype threat affects performance on inhibition tasks. We should note that neither this nor previous work argues that stereotype threat impairs participants' ability to inhibit via a top-down failure in cognitive control, as suggested in recent reviews (see Schmader et al., 2008). Instead, we argue that the increase in prepotent response tendency is due to potentiation (also see Ben-Zeev et al., 2005; Jamieson & Harkins, 2009; O'Brien & Crandall, 2003). In other words, motivation increases the likelihood of generating a prepotent response through a bottom-up mechanism (e.g., Lang & Bradley, 2008), rather than impairing top-down control. Consistent with this view, when the intervening task was framed as diagnostic, participants performed better, not worse, than when it was presented as an unrelated filler task. This result replicates Jamieson and Harkins's (2007) finding that, despite the fact that threatened females look in the wrong direction more often than controls on an antisaccade task, they performed better than control participants when they had time to implement correction. Likewise, McFall et al. (2009) found that, when given sufficient time to counter the prepotent tendency to read the word, participants subject to evaluation performed better on a Stroop task than participants who were not evaluated.

It is important to note that although the facilitation effect of threat on intervening task performance observed in this research can be explained by a single-process motivation model, additional processes may be at work in other stereotype threat situations. The possibility that multiple processes are involved in producing stereotype threat effects is consistent with the findings of previous studies that have made arguments for the contributions of cognitive (e.g., Schmader et al., 2008) and affective (e.g., Bosson, Haymovitz, & Pinel, 2004) factors, and with Steele, Spencer, and Aronson's (2002) argument that "stereotype threat effects are likely to be mediated in multiple ways—cognitively, affectively, and motivationally" (p. 397).

Given that females performed better on the intervening task when it was framed as diagnostic of math ability, what then accounts for the finding that participants who anticipate threat perform more poorly on nondiagnostic filler tasks than those who do not anticipate threat? The findings from Experiment 2 support a withdrawal of effort account, which is consistent with the pattern of results reported by Stone (2002) in his research on the effect of stereotype threat on self-handicapping. In that work, stigmatized participants practiced less than controls for an upcoming test that was diagnostic of their ability in a stereotyped domain. By practicing less, participants afforded themselves an excuse for their anticipated poor performance on the target task. Therefore, in both the current and that previous research, the anticipation of performing under conditions of stereotype threat led to a withdrawal of effort on an intervening task. However, threatened females in our filler task conditions were likely not engaging in self-handicapping, as they were in Stone's (2002) work, because the intervening task here was presented as unrelated to the stereotype.

Research on anticipatory coping suggests that the individuals seek to prepare themselves for likely or certain upcoming stressful events (e.g., Folkman & Lazarus, 1985). Thus, rather than self-handicapping, participants withdrawing effort in the current research could instead have been attempting to conserve resources for the upcoming task. This notion is consistent with recent work that demonstrates that individuals limit the amount of cognitive resources they devote to intervening tasks when they expect to need those resources to perform well in the future (Muraven, Shmueli, & Burkley, 2006). In the case of the current research, females anticipating a math test on which they know there are gender differences may seek to conserve cognitive resources on the unrelated filler task if those resources would help them perform well on the upcoming task.

Finding that a stereotype threat manipulation affects participants' performance differently depending on whether threat is anticipated or experienced in the present may not be surprising if one considers that the experience of stereotype threat is inherently motivating. That is, to be threatened by the prospect of confirming a negative stereotype, one must care about the fact that one's performance could reflect badly on the self (Inzlicht & Ben-Zeev, 2000) and/or one's group (Jamieson & Harkins, 2010; Wout, Danso, Jackson, & Spencer, 2008). Thus, if performance is measured on the task at hand, one might expect stigmatized individuals to be concerned with their performance so as to disconfirm the negative stereotype (e.g., Jamieson & Harkins, 2007, 2009; Kray et al., 2001). However, when performance is measured on filler tasks leading up to the threat-relevant situation, stigmatized individuals have no reason to be motivated. Filler task performance may instead index disengagement/withdrawal of effort (Keller, 2002; Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999).

In addition to informing us about mechanisms underlying stereotype threat effects, the current work has broad relevance for research in social psychology as well as for other psychological sciences. In recent years, there has been a shift from documenting the effects of psychological processes on outcome variables (performance, behavior, health, etc.) to identifying the mechanisms that mediate these effects. The intervening tasks method has been used as a method for

studying mechanisms (e.g., Ben-Zeev et al., 2005; Forbes et al., 2010; Maier et al., 2008; Schmader & Johns, 2003), but the current work suggests that researchers should exercise caution when drawing conclusions about mediating processes using this method because the processes that characterize performance on intervening tasks may differ from those that manifest on focal, diagnostic tasks.

As a result, we would argue that researchers should seek to measure underlying mechanisms during focal task performance whenever possible. In those cases where this is not feasible, converging evidence from a series of studies can be a powerful research tool for drawing conclusions about processes. For instance, Spencer, Zanna, and Fong (2005) argue that experimental-causal-chain designs, in which multiple studies examine a process as both an effect of a psychological state and a cause of the dependent variable, may be preferred over single experiments that make causal claims using statistical mediation analyses (also see Bullock, Green, & Ha, 2010).

In any case, the current research shows that the use of the intervening task method on its own may not be a valid method for measuring mechanisms because intervening tasks do not always capture the processes that are active during focal task performance. As a result, instead of helping us understand the mediating process in operation on the focal task, performance on the intervening task may simply present us with some other behavior to be explained, as was the case in the current research. Of course, if one is interested in studying anticipatory processes, this method may be quite useful. That is, measuring performance on unrelated filler tasks could be used to examine how anticipating a psychological state affects individuals' behavior, thoughts, and so on. For instance, a recent study explored the impact of expectations of cross-race interactions on reading comprehension (Sommers, Warp, & Mahoney, 2008). White participants who anticipated interacting with individuals of difference races exhibited facilitated performance on SAT reading comprehension problems, which prompted the authors to conclude that expectation of interacting with a racially diverse group can promote more thorough information processing. Thus, although the intervening task method is not ideal for measuring mediating mechanisms, it can provide insight into anticipatory processes.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Financial Disclosure/Funding

The authors received no financial support for the research and/or authorship of this article.

Notes

1. This pattern of results was produced when the target was displayed for 250 ms. When the target was displayed for only 150 ms, threatened participants performed worse, not better, than nonthreatened participants. Under these conditions, threatened participants did not have sufficient time to correct for their potentiated tendency to look toward the cue.
2. Inzlicht, McKay, and Aronson (2006) examined the effect of anticipated threat on an intervening task that was presented as unrelated to the stereotype. We, too, presented the intervening task as unrelated to the stereotype in the filler condition, which is warranted by the manipulation checks. However, typically when using blatant manipulations of stereotype threat, participants in the no threat control condition are specifically told that there are no group differences (e.g., Jamieson & Harkins, 2007). We also ran an additional no Stroop threat control condition in which participants were explicitly told that gender differences did not exist on the intervening Stroop task. Using this control condition produced the same pattern of findings as was observed in Experiment 1 when participants in the filler condition served as the control group.
3. The control list from Experiment 1 cannot be used to measure effort withdrawal in the filler condition because the focal task is performance on the incongruent list, which requires a greater degree of cognitive effort to perform than the control task. Thus, the amount of effort one put into the control list could be affected by the fact that one had just performed the more difficult, incongruent list, or anticipated doing so.

References

Ashcraft, M. H., & Kirk, E. P. (2001). The relationship between working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*, 224-237.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*, 174-181.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*, 247-255.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*, 626-633.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550-558.

Carr, P. B., & Steele, C. M. (in press). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19-23.

Folkman, S., & Lazarus, R. S. (1985). If it changes, it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, *48*, 150-170.

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology*, *99*, 740-754.

Harkins, S. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *91*, 436-455.

Hess, T. M., Hinson, J. T., & Hodges, E. A. (2009). Moderators of and mechanisms underlying stereotype threat effects on older adults' memory performance. *Experimental Aging Research*, *35*, 153-177.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365-371.

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigmas as ego depletion. *Psychological Science*, *17*, 262-269.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544-564.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin*, *35*, 1301-1314.

Jamieson, J. P., & Harkins, S. G. (2010). Evaluation is necessary to produce stereotype threat performance effects. *Social Influence*, *5*, 75-86.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169-183.

Karau, S. J., & Williams, S. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*, 681-706.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47*, 193-198.

Kirk, R. (1995). *Experimental design*. Pacific Grove, CA: Brooks/Cole.

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, *80*, 942-958.

Lang, P., & Bradley, M. (2008). Appetitive and defensive motivation is the substrate of emotion. In A. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 51-65). New York, NY: Psychology Press.

MacLeod, C. M. (1991). A half century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.

Maier, M. A., Elliot, A. J., & Lichtenfeld, S. (2008). Mediation of the negative effect of red on intellectual performance. *Personality and Social Psychology Bulletin*, *34*, 1530-1540.

McFall, S. R., Jamieson, J. P., & Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *96*, 135-154.

Muraven, M., Shmueli, D., & Burkley, E. (2006). Conserving self-control strength. *Journal of Personality and Social Psychology*, *91*, 524-537.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, *29*, 782-789.

Roberts, R. J., Hagar, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*, 374-393.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440-452.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336-356.

Seibt, B., & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*, 38-56.

Sommers, S. R., Warp, L. S., & Mahoney, C. C. (2008). Cognitive effects of racial diversity: White individuals' information processing in heterogeneous groups. *Journal of Experimental Social Psychology*, *44*, 1129-1136.

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.

Spencer, S. J., Zanna, M., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845-851.

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.

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. 23, pp. 379-440). New York, NY: Academic Press.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in White athletes. *Personality and Social Psychology Bulletin*, *28*, 1667-1678.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, *77*, 1213-1227.

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology*, *44*, 792-799.