Motivation Science
2016, Vol. 2, No. 2, 85–96

2333-8113/16/$12.00 http://dx.doi.org/10.1037/mot0000032

Threat Does Not Make the Mind Wander: Reconsidering the Effect of Stereotype Threat on Mind-Wandering

Adam J. Brown and Stephen G. Harkins
Northeastern University

Virtually all of the evidence suggesting that working memory deficits mediate the effect of stereotype threat on task performance relies on multiple task paradigms in which a seemingly unrelated task is performed after the manipulation of stereotype threat but before performance on the focal measure (e.g., Graduate Record Examination [GRE] math problems). While this intervening task is often a working memory measure, an indirect index of mind-wandering, the sustained attention to response task (SART), has also been used to argue that mind-wandering contributes to the working memory deficit. We tested the direct effect of stereotype threat on mind-wandering by describing the SART as either related to the gender-math stereotype or an unrelated attention task, while also manipulating stereotype threat for an anticipated GRE math test. We found that SART-stereotype threat participants outperformed participants who were told that the SART was an unrelated task, regardless of the stereotype threat condition to which they were assigned for the anticipated math test. These findings show that the assumption that performance on the intervening, unrelated task is debilitated by a working memory deficit is unwarranted and are, instead, consistent with the mere effort account.

Keywords: stereotype threat, mind-wandering, mere effort, intervening task, motivation

For a number of years, explanations for performance debilitation in social psychology have focused on cognitive processes, and, for the most part, have ignored the potential role of motivation. Research on stereotype threat, the concern that one feels when there is the potential to confirm a negative stereotype about one's group as self-relevant (Steele & Aronson, 1995), is no different. Two decades of research has demonstrated the deleterious effects of a number of stereotypes in performance situations (e.g., African Americans and academics; Steele & Aronson, 1995; women and mathematic ability; Spencer, Steele, & Quinn, 1999; individuals of lower SES and academic performance; Croizet & Claire, 1998). Although a number of processes have been identified as contributing to this performance debilitation (e.g., increased physiological stress response; Blascovich, Spencer, Quinn, & Steele, 2001), a growing consensus favors the position that these processes work by affecting working memory efficiency (Schmader, Hall, & Croft, 2015; Schmader, Johns, & Forbes, 2008).

Working Memory Deficit Account

According to the working memory deficit account, concern over confirming the stereotype produces thoughts and worries that take up cognitive resources that could be devoted to task performance, and it is this reduction in working memory capacity that produces the performance debilitation. To show that it is a reduction in working memory that produces the performance debilitation, researchers have used multiple task paradigms in which a seemingly unrelated task

This article was published Online First September 15, 2016.

Adam J. Brown and Stephen G. Harkins, Department of Psychology, Northeastern University.

This research was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences (Contract W5J9CQ-12-C-0046; Principal Investigator: Stephen G. Harkins). We thank David Desteno for his assistance with data analyses. The views, opinions, and/or findings contained in this report are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents.

Correspondence concerning this article should be addressed to Adam J. Brown, Department of Psychology, Northeastern University, 360 Huntington Avenue, 125 Nightingale Hall, Boston, MA 02115. E-mail: brown.ad@husky.neu.edu

is given between the manipulation of stereotype threat and the focal measure (e.g., Inzlicht, McKay, & Aronson, 2006; Mrazek et al., 2011; Rydell, Van Loo, & Boucher, 2014; Schmader & Johns, 2003). It is common practice to operationalize the intervening task as a working memory measure, performance on which is then used to demonstrate mediation of the effect of stereotype threat on task performance.

For example, in a recent review of the working memory deficit account, Schmader et al. (2015) cite the work of Mrazek et al. (2011) on mind-wandering as support for the notion that stereotype threat leads to deficits in working memory. Mind-wandering is defined as a decoupling of attention from the immediate task context toward unrelated concerns, and is suggested to be disruptive in a variety of performance situations (e.g., educational settings; Risko, Anderson, Sarwal, Engelhardt, & Kingstone, 2012; measures of attention; Cheyne, Solman, Carriere, & Smilek, 2009; reading comprehension; Smallwood, McSpadden, & Schooler, 2008).

In their research, Mrazek et al. (2011) told participants that they were going to complete two unrelated tasks. In the stereotype threat condition, female participants were seated between two male confederates at adjacent computer stations, the experimenter was male, and he told them that they would be taking a diagnostic math test designed to test for differences between men and women. First, however, the computer would administer a brief attention task. In the control condition, three female participants were run by a female experimenter, who described the upcoming test as a problem solving exercise preceded by the brief attention task.

All of the participants were then shown three difficult example problems like the ones that they were going to supposedly solve after the attention task. After this, participants completed the SART (Robertson, Manly, Andrade, Baddeley, & Yiend, 1997) on which participants are asked to respond to frequently appearing nontargets and to withhold responses for rare targets.

The SART is a go/no-go task that is used as an indirect measure of mind-wandering (Cheyne et al., 2009). The failure to withhold a response when the rare target appears, called a commission error, is the most commonly used, and thought of as the critical dependent measure derived from SART performance (e.g., Jackson & Balota, 2012; Manly, Roberston, Galloway, & Hawkins, 1999; Robertson et al., 1997; Seli, Cheyne, Barton, & Smilek, 2012). Mrazek et al. (2011) also report three additional dependent measures from the SART: omissions (i.e., the failure to respond on go trials), anticipations (i.e., responses that occur less than 100 ms after target onset), and the reaction time (RT) coefficient of variation (standard deviation of RT/mean RT).

Mrazek and his colleagues found that stereotype threat participants made more commission errors, more omissions, and exhibited larger RT coefficients of variation than the no threat controls. Mrazek et al. found no difference between conditions for anticipations. However, in research using the SART, it is not at all uncommon to find differences on some, but not all, of these measures (e.g., Carriere, Cheyne, Solman, & Smilek, 2010; Hart et al., 2012; Helton, Kern, & Walker, 2009) or for investigators to report only a subset of them (e.g., McVay & Kane, 2012; Mrazek, Smallwood, & Schooler, 2012). Taken together, these results led Mrazek et al. (2011) to conclude that stereotype threat leads to mind-wandering, which Schmader et al. (2015) cite as an example of how working memory suffers under stereotype threat, in this case because of an increase in task unrelated thoughts.

Two aspects of this work are noteworthy. First, in Mrazek et al.'s (2011) experiment,

> . . . the impaired performance associated with stereotype threat was observed on a task that itself was not pertinent to the stereotype. That is, there is no widely held stereotype that men and women differ in their performance on dull vigilance tasks, yet performance on this task was disrupted by the impending prospect of working on a separate task in a domain in which in which one's group is negatively stereotyped. (p. 1245)

Virtually all research aimed at showing that working memory deficits mediate the effect of stereotype threat on task performance uses this intervening (unrelated) task paradigm (Inzlicht et al., 2006; Mrazek et al., 2011; Rydell, McConnell, & Beilock, 2009; Rydell et al., 2014; Schmader & Johns, 2003). Second, according to Mrazek et al. (2011), this research serves as a test of an alternative, motivational account for the effect of stereotype threat on task performance, mere effort (Jamieson & Harkins, 2007), and the results do not support this account.

Mere Effort Account

The mere effort account argues that motivation, not deficits in working memory, is the core process producing stereotype threat performance effects. This account relies on Harkins's (2006) analysis of the effect of evaluation on task performance, which argues that the potential for evaluation motivates participants to want to perform well, potentiating the prepotent, or most likely, response on the given task. For example, when performing the Stroop Color-Word Test (Stroop, 1935), color words (e.g., green) are presented that are printed in some other color (e.g., red), and participants are asked to call out the color. However, the prepotent response is to read the word, which the mere effort account argues is potentiated by the potential for evaluation. If the prepotent response is correct, the mere effort account argues that the potential for evaluation facilitates performance. If incorrect and participants do not know it, or lack the knowledge or time required for correction, performance is debilitated. On the contrary, if participants recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated.

For example, on the Stroop, participants are well aware that the response of reading the word is incorrect, meaning that performance depends on the participants' motivation and ability to correct this response. McFall, Jamieson, and Harkins (2009) manipulated the ability to correct the prepotent response by manipulating the length of the response window for this task (1 s or 2 s). When participants were given a longer response window (2 s), participants subject to evaluation outperformed nonevaluated controls, but evaluated participants performed worse than nonevaluated participants when the response window was reduced (1 s). Harkins and his colleagues have also found support for the mere effort account of the effect of evaluation on performance on the remote associates task (Harkins, 2006), anagrams (McFall et al., 2009, Experiment 1), and the antisaccade task (McFall et al., 2009, Experiment 4).

The mere effort account of stereotype threat performance effects (Jamieson & Harkins, 2007) argues that stereotype threat operates like the potential for evaluation in that threat motivates stigmatized participants to want to perform well on the task. In fact, in this case the participant's performance will not only reflect on him or her, but also on his or her group. Jamieson and Harkins (2007) found support for the mere effort account using the antisaccade task (Hallett, 1978), an inhibition task like the Stroop. Support for the mere effort account of stereotype threat has also been found on other cognitive (e.g., GRE quantitative problems; Jamieson & Harkins, 2009; Stroop Color-Word Task; Jamieson & Harkins, 2011; Modular Arithmetic; Seitchik, & Harkins, 2015; Seitchik, Jamieson, & Harkins, 2014) and sensorimotor tasks (e.g., Huber, Brown, & Sternad, 2016; Huber, Seitchik, Brown, Sternad, & Harkins, 2015).

Mrazek et al. (2011) argue that the mere effort account cannot explain their participants' performance on the SART. They write: "the mere effort explanation [Jamieson & Harkins, 2007, 2009] might predict increased attention to the task as people exert more effort in an attempt to disconfirm the stereotype" (p. 1247). As a result, it would be expected that stereotype threat participants would outperform control participants. More specifically, on this task, the prepotent response should be to press the response key (90% of the stimuli require a keypress), and stereotype threat should potentiate this response. However, the heightened motivation of the stereotype threat participants should also prompt them to attempt to correct for this tendency, and the 2 s response window provides them with the time to do so (McFall et al., 2009). Thus, on the SART, with a 2 s response window, mere effort would predict that stereotype threat participants would outperform no threat controls.[1]

Instead, consistent with the working memory account, Mrazek et al. (2011) note that stereotype threat "produced a marked decrease in attention that subsequently impaired performance" (p. 1247). However, this analysis as-

[1] Importantly, the time course for responses to the Stroop and the SART are not the same. While 2 s is long enough for stereotype threat participants to enact the correction of the prepotent response on either task, the brief response window used in McFall et al. (2009) for the Stroop (1 s) may not be short enough to prevent correction on the SART. In fact, the average RT for a correct response on the SART is approximately 300 ms, suggesting that a response window of 400 ms or less would be needed to obtain the performance reversal found in McFall et al. (2009).

sumes that performance on the SART is debilitated by a working memory deficit just as a working memory deficit will debilitate performance on the following, focal GRE task, but this assumption is unwarranted.

Using the same intervening task paradigm as Mrazek et al. (2011), Inzlicht et al. (2006) found that threatened participants performed more poorly than control participants on an unrelated Stroop task that intervened between the stereotype threat manipulation and a GRE math test, even though the participants were provided enough time for correction on the Stroop. As a result, just as with Mrazek et al.'s SART outcome, this finding would appear to be inconsistent with mere effort and consistent with a working memory deficit account.

Instead, consistent with the mere effort account, Jamieson and Harkins (2011) have shown that when women were told that performance on the Stroop was related to the gender-math stereotype, they outperformed the women in the unrelated conditions, whether they were assigned to the stereotype threat or no threat condition for the anticipated math test. Thus, the direct effect of stereotype threat on Stroop performance was not the same as when it was an intervening task. In fact, in a second experiment, Jamieson and Harkins (2011) showed that stereotype threat participants withdrew effort on the Stroop when it was an intervening task to save themselves for performance on the stereotyped GRE task.

Current Research

If stereotype threat participants withdraw effort on the intervening task when it is not described as related to the gender-math stereotype, then it is possible that Mrazek et al. (2011) found poor SART performance (i.e., failures of controlled attention) for their stereotype threat participants because of differences in motivation rather than differences in working memory. To test this hypothesis, in the present research, we used Mrazek et al.'s (2011) procedures, but included conditions in which the SART was described as a math-related task on which there were gender differences.

In Mrazek et al.'s (2011) experiment, participants were given 2 s to provide their responses, providing ample time for correcting the prepotent response to press the response key. Given this, the mere effort account makes four predictions. First, participants in the GRE-stereotype threat/SART-unrelated condition will perform more poorly than GRE-no stereotype threat/SART-unrelated participants, replicating Mrazek et al.'s (2011) finding. As suggested by Jamieson and Harkins (2011), participants who are anticipating taking the GRE under stereotype threat will withdraw effort on the SART when it is described as an unrelated task. Second, in the SART-stereotype threat conditions, the difference in performance between GRE-stereotype threat and GRE-no stereotype threat participants should be eliminated, replicating Jamieson and Harkins (2011). That is, when performance on the SART is related to the gender-math stereotype, the participants in these two conditions should be equally and highly motivated to perform well on the SART, whether or not they anticipate taking the math test under stereotype threat.

In addition, these participants in the SART-stereotype threat conditions should outperform participants in the SART-unrelated conditions whether they are assigned to the GRE-no stereotype threat condition (Prediction 3) or to the GRE-stereotype threat condition (Prediction 4). With respect to the third prediction, the GRE-no stereotype threat/SART-unrelated participants have no reason to save effort for the upcoming GRE test because they have not been threatened, but they also are not highly motivated to perform well on the SART because it has been described only as an unrelated attention task. Thus, the SART-related participants who are highly motivated should outperform them. Of course, the fourth prediction follows from the other predictions in that the GRE-stereotype threat/SART-unrelated participants are predicted to perform more poorly than their GRE-no stereotype threat/SART-unrelated counterparts (Prediction 1), who, in turn, are predicted to perform more poorly than participants in the SART-related conditions (Prediction 3). Taken together, these predictions suggest that there will be three levels of performance in the paradigm (low, moderate, and high), which are depicted in Figure 1.

Of course, as in Mrazek et al. (2011), the working memory account predicts poorer performance in the GRE-stereotype threat/SART-unrelated condition than in the GRE-no stereotype threat/SART-unrelated condition. This account also predicts that participants in the

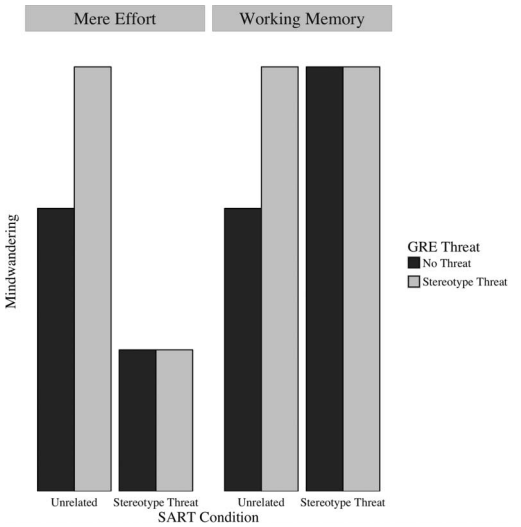

Figure 1. Predictions for the mere effort and working memory deficit accounts.

GRE-stereotype threat/SART-stereotype threat and GRE-no stereotype threat/SART-stereotype threat conditions will perform as poorly as the participants in the GRE-stereotype threat/SART-unrelated condition, and more poorly than participants in the GRE-no stereotype threat/SART-unrelated condition. After all, if SART performance is debilitated by the threat produced by the anticipated math test as argued by the working memory account, when performance on the SART itself is described as stereotyped, performance should also be debilitated. Thus, the working memory account would predict that participants assigned to stereotype threat conditions for either the anticipated math test, the SART, or both should perform more poorly than participants assigned to the GRE-no stereotype threat/SART-unrelated condition.[2] These predictions are also depicted in Figure 1.

Method

Participants

Seventy-three female undergraduate students participated in exchange for partial course credit. To determine the sample size the mean of the four effect sizes reported in Study 1 of Mrazek et al. (2011) was calculated, $d = .79$. Using G*power 3 software (Faul, Erdfelder, Lang, & Buchner, 2007), given this effect size, the minimum total sample size to achieve 80% power in the current design was $N = 53$. This value was used as a minimum sample size, but data collection continued until the end of the semester, yielding our sample.

Procedure

Participants were greeted by a male experimenter who told them that they would be completing a math test. In the stereotype threat condition, he went on to say:

> As you may know, there has been some controversy about whether there are gender differences in performance on math. Previous research has demonstrated that gender differences exist on some tests, but not on others. In our lab we examine performance on both kinds of tests. The test that you are about to take has been shown to produce gender differences.

In the no stereotype threat condition, the only difference was that the experimenter said that the "test that you are about to take **has not** been shown to produce gender differences." This manipulation has been used successfully in past research (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2009; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999).

The experimenter then gave the participants three sample GRE problems to look over. In the Mrazek et al. (2011) replication conditions, crossed with the GRE-stereotype threat manipulation, half of the participants were told that prior to beginning the math test, they would be completing a short, unrelated attention task. The other half of the participants were informed that they would be completing the SART, which was an attention measure that was closely related to math ability and that this measure produced gender differences. Participants then completed a 240 trial SART, with 216 nontarget trials (O) and 24 target trials (Q). Stimuli were presented for 2 s each with a 500 ms interstimulus interval (Cheyne et al., 2009). Participants

[2] Alternatively, this account could predict that participants in the GRE-stereotype threat/SART stereotype threat condition would show even more mind wandering than participants who only hear the stereotype threat manipulation for one of these tasks. The predictions made for the working memory deficit account describe what should happen, at a minimum, according to this explanation.

were instructed to press the spacebar as quickly as possible when a nontarget (O) appeared and to refrain from responding when the target appeared (Q).

Four measures of mind-wandering were derived from SART performance. Trials on which participants made a keypress when the rare target appeared (i.e., trials that required participants to refrain from responding) were coded as commission errors. Keypresses on trials that required participants to respond that were faster than 100 ms were coded as anticipations. Failures to respond on trials that required a keypress were coded as omission errors. The RT coefficient of variation was calculated by dividing the standard deviation of RTs by the mean of RTs for all trials that required a keypress that were not coded as anticipations or omissions.

After finishing the SART, participants completed two manipulation check questionnaires, one about the SART and one about the anticipated math test. Participants responded to two critical items on each form asking, "To what extent do you believe that gender differences exist on this task?" (1 = *no gender difference* and 11 = *gender differences*) and "Who do you believe performs better on this task?" (1 = *males perform better*, 6 = *no gender difference*, and 11 = *females perform better*). Participants were also asked, "To what extent is performance on this task related to math ability?" (1 = *not at all* and 11 = *related*).

The experiment concluded after the two questionnaires were completed and no math test was given as in Mrazek et al. (2011). No participants were excluded from any analyses.

Results

Manipulation Checks

Manipulation checks were analyzed using a 2 (anticipated threat: GRE-stereotype threat vs. GRE-no stereotype threat) × 2 (SART description: SART-stereotype threat vs. SART-unrelated) between-subjects analysis of variance (ANOVA).

SART manipulation check. SART-stereotype threat participants believed that this task was related to math ability to a greater extent ($M = 7.67$, $SD = 2.99$) than SART-unrelated participants ($M = 5.62$, $SD = 2.60$), $F(1, 69) = 9.58$, $p < .01$, $\eta_p^2 = .122$. SART-stereotype threat participants also reported that gender differences existed on the SART to a greater extent ($M = 6.61$, $SD = 2.35$) than SART-unrelated participants ($M = 3.70$, $SD = 2.77$), $F(1, 69) = 22.73$, $p < .001$, $\eta_p^2 = .248$. Finally, SART-stereotype threat participants reported that men performed better than women on the SART ($M = 4.56$, $SD = 1.63$) to a greater extent than SART-unrelated participants ($M = 5.70$, $SD = 1.58$), $F(1, 69) = 9.05$, $p < .01$, $\eta_p^2 = .116$. The effect of anticipated threat and the interaction did not approach significance in any analysis (Fs < 1).

Math test manipulation check. Participants did not differ in their ratings of the extent to which performance on the anticipated math test reflected math ability, Fs < 1.4, ps $> .24$ (overall $M = 9.07$). GRE-stereotype threat participants reported that the math test would produce gender differences to a greater extent ($M = 6.61$, $SD = 2.89$) than GRE-no stereotype threat participants ($M = 3.65$, $SD = 3.08$), $F(1, 69) = 17.60$, $p < .001$, $\eta_p^2 = .203$. GRE-stereotype threat participants also reported that men would outperform women on the upcoming math test to a greater extent ($M = 4.39$, $SD = 1.74$) than GRE-no stereotype threat participants ($M = 5.59$, $SD = 1.34$), $F(1, 69) = 10.71$, $p < .01$, $\eta_p^2 = .134$. The effect of SART description and the interaction did not approach significance in either analysis (Fs < 1.08). Taken together, these results indicate that our manipulations were successful.

SART Performance

Statistical analysis of performance. Given the specific predictions made to test the mere effort and the working memory deficit accounts, linear contrasts were used in a multivariate ANOVA (MANOVA) with all four measures of mind-wandering as the dependent variables (i.e., commission errors, omissions, anticipations, and RT coefficient of variation). The four predictions from the mere effort account can be tested in a linear contrast that predicts that the performance of participants in the SART-stereotype threat condition will not differ as a function of whether they anticipated taking the math test under stereotype threat or not (contrast weights: −3, −3) and will exhibit less mind-wandering than participants in the GRE-no stereotype threat/SART-unrelated condition (con-

trast weight: 1). These three groups are also predicted to show less mind-wandering than participants in the GRE-stereotype threat/SART-unrelated condition (contrast weight: 5). These weights were chosen to create three levels of performance (i.e., low, moderate, high), in line with the predictions depicted in Figure 1.

The working memory account predicts that participants given the stereotype threat manipulation for the anticipated math test, the SART, or both will exhibit equivalent amounts of mind-wandering (contrast weights: 1, 1, 1), all more than the mind-wandering found in the GRE-no stereotype threat/SART-unrelated condition (contrast weight: −3). Again, these weights map onto those depicted in Figure 1, such that the working memory account suggests that being assigned to a stereotype threat condition for either measure should lead to debilitated performance on the SART.

Mere effort. A MANOVA using the contrast weights described above was conducted to test the predictions of the mere effort account for the four dependent measures derived from the SART. There was a significant overall effect for the mere effort contrast, Wilk's $\Lambda = .657$, $F(4, 66) = 8.63$, $p < .001$, $\eta_p^2 = .343$. The univariate tests of these predictions also yielded significant effects supporting the mere effort account for commission errors, $F(1, 69) = 28.78$, $p < .001$ $\eta_p^2 = .294$, omissions, $F(1, 69) = 4.91$, $p = .030$, $\eta_p^2 = .066$, anticipations, $F(1, 69) = 11.42$, $p < .01$, $\eta_p^2 = .142$, and RT coefficient of variation, $F(1, 69) = 11.75$, $p = .001$, $\eta_p^2 = .146$. Additionally, the proportion of variance in the $SS_{between}$ that was explained by this contrast was calculated (Rosnow, Rosenthal, & Rubin, 2000). For commission errors, the contrast testing the mere effort account predictions accounted for 98.7% of the variance between the conditions. For omissions, this contrast accounted for 75.8% of the variance. The mere effort predictions accounted for 71.8% of the variance for anticipations. Finally, 99.2% of the variance was captured by the contrast for RT coefficient of variation. The residuals of this contrast were also tested to determine whether any systematic variance remained for commission errors, $F(1, 69) < 1$, omissions, $F(1, 69) = 1.59$, $p = .212$, $\eta_p^2 = .023$, anticipations, $F(1, 69) = 4.55$, $p = .037$, $\eta_p^2 = .062$, and RT coefficient of variation, $F(1, 69) < 1$. Means and standard deviations for all four dependent variables are presented in Table 1.[3]

Working memory. A second MANOVA was conducted using the contrast weights that tested the working memory deficit predictions: experiencing stereotype threat for either the anticipated math test, the SART, or both, leads to equivalent performance, which is worse than that found in the GRE-no threat/SART-unrelated condition. The multivariate test of this prediction was not significant, Wilk's $\Lambda = .993$, $F(4, 66) = 1.15$, $p = .33$, $\eta_p^2 = .065$. None of the univariate tests of this prediction were significant, $ps > .132$. The proportion of variance accounted for by the contrast was also calculated for these predictions. For commission errors, omissions, anticipations, and RT coefficient of variation, the contrast for the working memory deficit account predictions accounted for 0.4%, 11.5%, 14.5%, and 6.4% of the variance between groups, respectively. The residuals of this contrast were also tested to determine whether any systematic variance remained for commission errors, $F(1, 69) = 28.99$, $p < .001$, $\eta_p^2 = .296$, omissions, $F(1, 69) = 5.75$, $p = .019$, $\eta_p^2 = .077$, anticipations, $F(1, 69) = 13.64$, $p < .001$, $\eta_p^2 = .165$, and RT coefficient of variation, $F(1, 69) = 10.99$, $p < .01$, $\eta_p^2 = .139$.

Discussion

In a recent review of the working memory deficit account, Schmader et al. (2015) cite Mrazek et al.'s (2011) work on mind-wandering as evidence that difficulties in attentional focus produced by stereotype threat contribute to the working memory deficit. Indeed, Mrazek et al. (2011) found that when the SART was introduced as an unrelated attention task, anticipating a threatening math test (i.e., undergoing stereotype threat) led to more SART errors, higher RT coefficient of variation values, and more omissions, which, taken together, were interpreted as showing more mind-wandering than found for no stereotype threat controls. This pattern of findings is "consistent with a framework in which cognitive resources are depleted but is difficult to explain if threat elicits

[3] The means for omissions suggest a potential floor effect for this measure. However, all of the omissions occur in the GRE-stereotype threat/SART-stereotype threat condition, which was predicted to perform the worst.

Table 1
Means and Standard Deviations of the SART Measures by Condition

SART measure	GRE-no threat/SART unrelated (n = 19)	GRE-stereotype threat/SART unrelated (n = 18)	GRE-no threat/SART stereotype threat (n = 18)	GRE-stereotype threat/SART stereotype threat (n = 18)
Commission errors	7.11 (3.38)	10.44 (4.55)	5.00 (2.97)	4.89 (3.03)
Omissions	0 (0)	.33 (.97)	0 (0)	0 (0)
Anticipations	.37 (.68)	4.44 (7.09)	.55 (1.42)	.56 (1.10)
Reaction time Coefficient of variation (ms)	305 (105)	341 (97)	254 (72)	258 (76)

Note. SART = sustained attention to response task; GRE = Graduate Record Examination.

only greater effort and (presumably) attention toward the task (Mrazek et al., 2011)" (Schmader et al., 2015, p. 459).

In the current work, we replicated Mrazek et al.'s (2011) general pattern of findings on the SART with only minor differences. However, we also found that across all four SART measures mere effort predicted substantially more variance than the working memory account. Furthermore, nonsignificant tests of the mere effort contrast residuals suggest that these predictions accounted for all of the systematic variance in the means for three of the four SART measures, including commission errors (i.e., the critical measure of mind-wandering), whereas the tests of the working memory contrast residuals suggest that the data do not conform to the predictions drawn from that perspective.

This pattern of means suggests that when performance on the SART is introduced as unrelated to the manipulated stereotype, participants in the GRE-stereotype threat condition perform more poorly than their GRE-no stereotype threat counterparts. However, describing the SART as related to the same stereotype that was manipulated for the math test eliminated the difference between these two conditions. Critically, in line with the suggestion that undergoing stereotype threat leads to an increase in motivation, participants in the SART-stereotype threat conditions also outperformed participants in each of the SART-unrelated conditions.

Although these findings do not support Mrazek et al.'s (2011) explanation, an alternative working memory account could suggest that when the intervening task is described as unrelated to the stereotype, participants who expect to take the math test under threat are plagued by negative thoughts, which accounts for their poor performance on the mind-wandering task. On the contrary, when performance on the mind-wandering task is related to the stereotype, stereotype threat participants could be motivated to disconfirm the stereotype, and because the task is "easy," they can overcome the effects of the negative thoughts and outperform the SART-unrelated participants.

However, Jamieson and Harkins (2011, Exp. 2) have provided evidence that is inconsistent with this possibility. To do so, they used the prosaccade task (Roberts, Hagar, & Heron, 1994), performance on which has been shown to be positively related to the amount of effort exerted by participants but unaffected by deficits in processing capacity (Jamieson & Harkins, 2007, Exps. 3 and 4; Roberts et al., 1994). In their research, Jamieson and Harkins (2011, Exp. 2) crossed the framing of the intervening task (prosaccade threat vs. unrelated) and anticipated threat on the math test. Replicating our findings on the SART, women informed that gender differences existed on the intervening prosaccade task outperformed those assigned to the unrelated condition, regardless of anticipated threat condition. However, when the intervening task was described as unrelated to math ability, women who anticipated taking the GRE task under threat performed the same, simple prosaccade task more poorly than women not anticipating threat. This reduction in performance on the prosaccade task is inconsistent with a working memory account because a reduction in working memory capacity should not affect performance on this particular task. However, it is consistent with the argument that participants who are anticipating threat withdraw effort as they await their performance on the upcoming GRE test. Importantly, although these findings on the prosaccade task rule out

the working memory deficit plus motivation interpretation of the current results, they are uninformative concerning the effect of stereotype threat on mind-wandering behavior.

Our findings also suggest an alternative interpretation for research that both precedes and follows Mrazek et al.'s (2011) work. For example, Schmader et al. (2003) were the first to use the intervening task methodology to argue that stereotype threat produces a working memory deficit based on the fact that performance on an intervening working memory task mediated the effect of stereotype threat on task performance. Of course, instead of arguing that variation in the amount of negative thoughts produced differences in task performance, one can just as easily argue that the more motivated one was to perform well on the focal GRE test, the more one withdrew effort on the intervening task.

Subsequent to Mrazek et al.'s work, Rydell et al. (2014) used the intervening task paradigm to determine which working memory process(es) (inhibition, updating, shifting) was (were) responsible for threat effects. The present work suggests that in each of their experiments, participants who anticipated performing under threat simply withdrew effort on the intervening task, especially since we know that threatened participants perform better, not worse, than control participants on inhibition tasks (e.g., antisaccade task; Jamieson & Harkins, 2007; Stroop; Jamieson & Harkins, 2011) when they have time to correct for the prepotent response, as was the case in Rydell et al.'s experiments. The present work emphasizes the dangers inherent in using the intervening task paradigm to assess mediation, which is particularly problematic in the stereotype threat literature (e.g., Inzlicht et al., 2006; Mrazek et al., 2011; Rydell, McConnell, & Beilock, 2009; Rydell et al., 2014; Schmader & Johns, 2003).

The present work focused on Mrazek et al.'s (2011) first experiment, but they also conducted a second experiment in which they used what they described as a more direct indicator of mind-wandering. Women worked on a GRE test after a stereotype threat manipulation, and, as they performed, they responded to random probes asking them to rate the extent to which their minds were on the task. Threatened women reported more off-task thoughts, and this measure mediated the effect of stereotype threat on GRE performance. However, this probe methodology is problematic. The women in the stereotype threat condition were threatened, and then they were provided with a ready-made excuse for their potential poor performance, as in self-handicapping research (Ben-Zeev, Fein, & Inzlicht, 2005; Steele & Aronson, 1995; Stone, 2002). Thus, the more stereotype threat women were motivated to perform well, the more likely they may have been to take advantage of the opportunity to claim that they were bothered by off-task thoughts. As a result, these findings could be consistent with either account.

Similar to the current work, Inzlicht and Schmeichel (2012) suggest that motivation may also play an important, but largely unrecognized, role in accounting for ego depletion effects. They argue that after completing an onerous, initial self-control task, participants may feel that they deserve a break or to be rewarded for their hard work, leading them to exert less effort, or to pursue the more rewarding alternative on the second task. That is, it is not that there is some finite pool of cognitive resources that has been depleted, and is not available for the second task, but rather that the participants make strategic decisions about how to marshal their efforts on the second task, given their effortful performance on the first one. Consistent with this view, when "depleted" participants are offered incentives for performance on the second task, it turns out that they are not depleted (e.g., Muraven, Shmueli, & Burkley, 2006; Muraven & Slessareva, 2003). Thus, whereas our research suggests that threatened participants may withdraw effort on an intervening task to save themselves for the focal math test, Inzlicht and Schmeichel argue that participants may withdraw effort on a second task after putting a good effort into the first one. In both cases, poor performance that has been attributed to a cognitive deficit may instead reflect a reduction in motivation.

At a general level, the present work, along with that of Inzlicht and Schmeichel (2012), represents part of a movement toward once again integrating motivation into our accounts of task performance. More specifically, the present research supports the motivational, mere effort account for stereotype threat performance effects over the cognitive, working memory deficit account. Knowing what causes stereotype threat performance effects is impor-

tant because this knowledge is required for us to design interventions to counter these effects. For example, other performance findings suggest that a simple instruction is sufficient to eliminate the debilitation produced by stereotype threat (Jamieson & Harkins, 2009; Seitchik & Harkins, 2015). These interventions would not follow from a working memory account, because that approach argues that it is not that people do not know what to do; it is that they do not have the capacity to do it. The current and past research suggests otherwise, offering new approaches to improve performance in stigmatized individuals.

References

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41,* 174–181. http://dx.doi.org/10.1016/j.jesp.2003.11.007

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12,* 225–229. http://dx.doi.org/10.1111/1467-9280.00340

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. http://dx.doi.org/10.1016/S0022-1031(03)00039-8

Carriere, J. S. A., Cheyne, J. A., Solman, G. J., & Smilek, D. (2010). Age trends for failures of sustained attention. *Psychology and Aging, 25,* 569–574. http://dx.doi.org/10.1037/a0019363

Cheyne, A. J., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition, 111,* 98–113. http://dx.doi.org/10.1016/j.cognition.2008.12.009

Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype 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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296. http://dx.doi.org/10.1016/0042-6989(78)90218-3

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436–455. http://dx.doi.org/10.1037/0022-3514.91.3.436

Hart, E. P., Dumas, E. M., Reijntjes, R. H. A. M., van der Hiele, K., van den Bogaard, S. J. A., Middelkoop, H. A. M., . . . van Dijk, J. G. (2012). Deficient sustained attention to response task and P300 characteristics in early Huntington's disease. *Journal of Neurology, 259,* 1191–1198. http://dx.doi.org/10.1007/s00415-011-6334-0

Helton, W. S., Kern, R. P., & Walker, D. R. (2009). Conscious thought and the sustained attention to response task. *Consciousness and Cognition: An International Journal, 18,* 600–607. http://dx.doi.org/10.1016/j.concog.2009.06.002

Huber, M. E., Brown, A. J., & Sternad, D. (2016). Girls *can* play ball: Stereotype threat reduces variability in a motor skill. *Acta Psychologica, 169,* 79–87. http://dx.doi.org/10.1016/j.actpsy.2016.05.010

Huber, M. E., Seitchik, A. E., Brown, A. J., Sternad, D., & Harkins, S. G. (2015). The effect of stereotype threat on performance of a rhythmic motor skill. *Journal of Experimental Psychology: Human Perception and Performance, 41,* 525–541. http://dx.doi.org/10.1037/xhp0000039

Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science, 17,* 262–269. http://dx.doi.org/10.1111/j.1467-9280.2006.01695.x

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7,* 450–463. http://dx.doi.org/10.1177/1745691612454134

Jackson, J. D., & Balota, D. A. (2012). Mind-wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging, 27,* 106–119. http://dx.doi.org/10.1037/a0023933

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93,* 544–564. http://dx.doi.org/10.1037/0022-3514.93.4.544

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin, 35,* 1301–1314. http://dx.doi.org/10.1177/0146167209335165

Jamieson, J. P., & Harkins, S. G. (2011). The intervening task method: Implications for measuring mediation. *Personality and Social Psychology Bul-*

letin, 37, 652–661. http://dx.doi.org/10.1177/0146167211399776

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29,* 371–381. http://dx.doi.org/10.1177/0146167202250218

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia, 37,* 661–670. http://dx.doi.org/10.1016/S0028-3932(98)00127-4

McFall, S., 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. http://dx.doi.org/10.1037/a0012878

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141,* 302–320. http://dx.doi.org/10.1037/a0025250

Mrazek, M. D., Chin, J. M., Schmader, T., Hartson, K. A., Smallwood, J., & Schooler, J. W. (2011). Threatened to distraction: Mind-wandering as a consequence of stereotype threat. *Journal of Experimental Social Psychology, 47,* 1243–1248. http://dx.doi.org/10.1016/j.jesp.2011.05.011

Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2012). Mindfulness and mind-wandering: Finding convergence through opposing constructs. *Emotion, 12,* 442–448. http://dx.doi.org/10.1037/a0026678

Muraven, M., Shmueli, D., & Burkley, E. (2006). Conserving self-control strength. *Journal of Personality and Social Psychology, 91,* 524–537. http://dx.doi.org/10.1037/0022-3514.91.3.524

Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29,* 894–906. http://dx.doi.org/10.1177/0146167203029007008

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. http://dx.doi.org/10.1177/0146167203029006010

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393. http://dx.doi.org/10.1037/0096-3445.123.4.374

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology, 26,* 234–242. http://dx.doi.org/10.1002/acp.1814

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia, 35,* 747–758. http://dx.doi.org/10.1016/S0028-3932(97)00015-8

Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science, 11,* 446–453. http://dx.doi.org/10.1111/1467-9280.00287

Rydell, B. 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

Rydell, R. J., Van Loo, K. J., & Boucher, K. L. (2014). Stereotype threat and executive functions: Which functions mediate different threat-related outcomes? *Personality and Social Psychology Bulletin, 40,* 377–390. http://dx.doi.org/10.1177/0146167213513475

Schmader, T., Hall, W., & Croft, A. (2015). Stereotype threat in intergroup relations. In M. Mikulincer & P. R. Shaver (Eds.), *APA handbook of personality and social psychology* (Vol. 2, pp. 447–471). Washington, DC: American Psychological Association.

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

Seitchik, A. E., Brown, A. J., & Harkins, S. G. (in press). Social facilitation: using the molecular to inform the molar. In S. G. Harkins, K. D. Williams, & J. Burger (Eds.), *The Oxford handbook of social influence*. New York, NY: Oxford University Press.

Seitchik, A. E., & Harkins, S. G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. *Journal of Experimental Social Psychology, 61,* 19–30. http://dx.doi.org/10.1016/j.jesp.2015.06.006

Seitchik, A. E., Jamieson, J., & Harkins, S. G. (2014). Reading between the lines: Subtle stereotype threat cues can motivate performance. *Social Influence, 9,* 52–68. http://dx.doi.org/10.1080/15534510.2012.746206

Seli, P., Cheyne, J. A., Barton, K. R., & Smilek, D. (2012). Consistency of sustained attention across

modalities: Comparing visual and auditory versions of the SART. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66,* 44–50. http://dx.doi.org/10.1037/a0025111

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36,* 1144–1150. http://dx.doi.org/10.3758/MC.36.6.1144

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., & 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

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. http://dx.doi.org/10.1177/014616702237648

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. http://dx.doi.org/10.1037/h0054651

Received March 1, 2016
Revision received August 10, 2016
Accepted August 17, 2016 ■