Journal of Experimental Social Psychology 61 (2015) 19–30

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Stereotype threat, mental arithmetic, and the mere effort account

Allison E. Seitchik *, Stephen G. Harkins

Northeastern University, Department of Psychology, 360 Huntington Avenue, 125 Nightingale Hall, Boston, MA 02115, USA

HIGHLIGHTS

• Pitted working memory against a motivational account for stereotype threat effects
• The test bed was horizontal subtraction problems solved in the participants' heads.
• The motivational, mere effort, account was supported.

ARTICLE INFO

Article history:
Received 8 January 2014
Revised 26 May 2015
Accepted 22 June 2015
Available online 30 June 2015

Keywords:
Stereotype threat
Motivation
Mental arithmetic
Performance
Working memory

ABSTRACT

The currently prevailing explanation for stereotype threat (ST) debilitation effects argues for working memory interference as the proximal mediator. Using mental arithmetic problems as the test bed, Beilock et al. (2007) have spelled out in greater detail exactly how this process might work. They propose that worries resulting from activation of the negative stereotype occupy the phonological loop, taking up capacity that could be used to remember the intermediate values produced when solving horizontal mental subtraction problems. In the current work, we test an alternative, motivational account for this effect, mere effort. The mere effort account argues that ST motivates stigmatized participants to want to perform well, potentiating the prepotent response on the given task. In Experiment 1 we identified a possible prepotent response for horizontal subtraction, termed the method of adjustment (e.g., adjust the second number to the nearest 10, subtract the two numbers, and then add the adjustment). Consistent with the mere effort account, Experiment 2 showed that ST potentiated the prepotent approach, the method of adjustment. Experiment 3 pitted the mere effort account against the working memory account. Working memory predicts debilitation effects on horizontal subtraction problems when participants need to use the phonological loop (i.e., entering answers from left-to-right), whereas mere effort predicts that the potentiated use of the method of adjustment should facilitate performance when answers must be entered from left-to-right. Results supported the mere effort account. Finally, Experiment 4 showed that when we control for the effect of the potentiated prepotent response, instead of performing more poorly, threatened participants perform better than no threat participants. Overall, these experiments support the mere effort account, which argues for motivation as a core process in producing ST effects.

1. Introduction

Stereotype threat (ST) refers to the concern that is experienced when one feels "at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995, abstract). ST has been shown to negatively impact the performance of stigmatized individuals in a variety of domains (e.g., African-American's underperformance on standardized tests (e.g., Steele & Aronson, 1995); White males' athletic inferiority (e.g., Stone, 2002); and females' lack of ability in math and science domains (e.g., Ben-Zeev, Fein, & Inzlicht, 2005)). Although a number of processes have been identified as contributing to this performance debilitation, a growing consensus favors the position that these processes work by affecting working memory efficiency (e.g., Schmader, Hall, & Croft, 2015; Schmader, Johns, & Forbes, 2008).

1.1. Working memory account

At the core of the working memory account is the contention that cognitive resources that could be devoted to task performance are instead expended on processing the information resulting from the activation of the negative stereotype (e.g., Schmader et al., 2008). Using Baddeley's (e.g., Baddeley, 1986, 2000) multicomponent model of working memory, Beilock, Rydell, and McConnell (2007) have provided the most detailed account available of the process through which ST produces working memory interference. Baddeley proposes

* Corresponding author at: Harvard University, Department of Psychology, 33 Kirkland Street, Cambridge, MA 02138, USA.
E-mail addresses: seitchik.a@husky.neu.edu, aseitchik@g.harvard.edu (A.E. Seitchik), s.harkins@neu.edu (S.G. Harkins).

http://dx.doi.org/10.1016/j.jesp.2015.06.006
0022-1031/

four subsystems that temporarily store and process different types of information: (1) a limited-capacity central executive; (2) the phonological loop, which rehearses verbal information; (3) a visuospatial sketchpad, which serves as a mental blackboard and temporarily stores visual and spatial information; and (4) a multimodal episodic buffer, which connects information from the phonological loop, visuospatial sketchpad, and long-term memory into one episodic representation. Beilock et al. argue that although the pressure-induced worries produced by ST have some impact on central executive resources, these worries primarily occupy the phonological loop, which is thought to support inner speech (Beilock et al.).

Beilock et al. (2007) tested this hypothesis by examining performance on two types of math problems that rely equally on the central executive, but differ in their reliance on the phonological loop. Trbovich and LeFevre (2003) have shown that participants presented mental arithmetic problems in a horizontal format tend to rely on the phonological loop (remembering intermediate values in verbal form), whereas participants presented problems in a vertical format tend to rely on the visuospatial sketchpad (using a spatial mental work space in the same way that such problems are solved on paper). As a result, Beilock et al. argue that performance on horizontal problems, which requires the use of the phonological loop, should be disrupted by ST, whereas performance on vertical problems, which depends more on the visuospatial sketchpad, should not.

Beilock et al. (2007, Experiment 3A) tested this prediction by asking females under ST or no ST to complete modular arithmetic problems presented in a horizontal or a vertical orientation. For example, participants saw: $52 = 24 \pmod 3$. They were to subtract the second number from the first, divide by the mod, and then report, true or false, whether the result was a whole number.[1] Consistent with their hypothesis, Beilock et al. found that, under ST, on horizontal/high-demand problems (double-digit borrow subtraction), females performed significantly worse than their baseline (no threat) performance. There was no performance debilitation in any other condition. This pattern of finding supports Beilock et al.'s working memory deficit account, which suggests that threat produces worries that occupy the phonological loop, debilitating performance on horizontal subtraction problems that require the phonological loop to retain intermediate values.

1.2. Mere effort account

However, Jamieson and Harkins (2007) have proposed an alternative to this working memory account, mere effort, which argues that motivation, not deficits in working memory, is the core process producing ST 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, 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, or lack the knowledge or time required for correction, performance is debilitated. However, if participants recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated (cf. Zajonc, 1965).[2] Harkins and his colleagues have found support for the mere effort account of the effect of evaluation on performance on the Remote Associates Task (Harkins, 2006), anagrams (McFall, Jamieson & Harkins, 2009, Experiment 1), the Stroop task (McFall, Jamieson, & Harkins, 2009, Experiments 2 and 3), and the antisaccade task (McFall et al., 2009, Experiment 4).

The mere effort account of ST performance effects argues that ST operates like the potential for evaluation in that threat will motivate 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. To test the mere effort account, Jamieson and Harkins (2007) used the antisaccade task (Hallett, 1978), an inhibition task like the Stroop. This task requires participants to respond to a target presented on either side of a display screen after a cue appears on the opposite side of the display. The objective of the task is to accurately report the orientation of the target arrow (i.e., whether it is pointing up, left, or right). Participants are told not to look at the cue, but instead to look at the opposite side of the display where the target will be presented. However, participants have a reflexive-like prepotent tendency (i.e., reflexive saccade) to look at the cue that must be inhibited or corrected in order to perform well.

This task was framed as a measure of visuospatial capacity that was highly related to ability in math, allowing it to be described as producing (ST) or not producing (no ST) gender differences. Consistent with the mere effort account, on this task, when participants were given insufficient time to correct for the prepotent tendency (i.e., at a brief display time; e.g., 150 ms), the ST participants were less accurate in their identification of target orientation than controls, but when the display time was increased enough to allow enough time for correction (e.g., 250 ms), ST participants outperformed the controls (Jamieson & Harkins, 2007, Experiments 1 and 2).

Jamieson and Harkins (2007, Experiment 3) also used an eye tracker to conduct a more fine-grained analysis of performance on this antisaccade task at a display time that permitted correction. Consistent with the mere effort account, Jamieson and Harkins (2007) found that ST participants launched more reflexive saccades than controls (i.e., looked the wrong direction at the cue more often). However, they were also faster to launch corrective saccades (eye movements back to the target) than controls, as well as to respond more quickly to the target once their eyes arrived at the target site. As a result, overall, ST participants outperformed the controls.

In addition to the antisaccade trials, participants performed a prosaccade task in which the cue appears on the same side as the target. In this case, the prepotent response (to look at the cue) is the correct response, and all participants tend to make this response whether they are in the ST condition or not (i.e., the prepotent response is at a ceiling). Despite the fact that the participants' eyes arrive at the target site at essentially the same time, the threat participants were faster than no threat participants to press the response button as a result of their greater motivation to perform well (Jamieson & Harkins, 2007). The mere effort account of ST effects has also been supported using other tasks (e.g., Graduate Record Exam quantitative problems, Jamieson & Harkins, 2009; Seitchik, Jamieson, & Harkins, 2014; Stroop Color-Word Task, Jamieson & Harkins, 2011).

[1] As Beilock (2010) has noted, there are other ways to solve these modular arithmetic problems. For example, the numbers to be subtracted can each be divided by the mod. If the remainders are the same, the answer is true; if not, the answer is false. In the research to be described, we used the same instructions as those used by Beilock et al. (2007).

[2] Drive theory (Zajonc, 1965), like the mere effort account, accords a central role to prepotent or dominant responses. It argues that the presence of others produces arousal, which increases drive. Increased drive enhances the probability of the emission of dominant responses, which are likely to be correct on simple tasks but incorrect on difficult ones. In fact, Cottrell (1968, 1972) argued that this drive was the result of the participants' apprehension about the fact that they could be evaluated. However, in the case of mere effort, this potentiation results from the motivation to perform well, which should also lead to an effort to correct an incorrect response if the participant recognizes that his or her response is incorrect, knows the correct response, and has the opportunity to make it. In contrast, Cottrell's (1968) modification of Zajonc's (1965) drive theory suggests only that the positive or negative anticipation produced by the presence of others nonselectively energizes performance (i.e., potentiates the dominant response).

1.3. Current research

Previous work on the mere effort account (e.g., Jamieson & Harkins, 2007, 2009) has been cited as showing that motivation can contribute to ST performance effects (e.g., Inzlicht & Kang, 2010; Schmader et al., 2008, 2015; Shapiro & Williams, 2012), but not that it represents an alternative account. For example, Schmader et al. (2008) argue that, in the antisaccade paradigm, a working memory deficit accounts for the ST participants' inability to inhibit the tendency to look the wrong way, but motivation could account for the finding that ST participants launch corrective saccades faster than controls. In the current work, we pitted Beilock et al.'s (2007) elaborated working memory interference explanation against the mere effort account in a paradigm in which they make opposing predictions. We argue that the resulting findings show that mere effort can provide an account of ST performance effects on this task, whereas the working memory interference account cannot.

As a first step, we sought to determine whether mere effort could provide an account of the effect of ST on the performance of Beilock et al.'s (2007) horizontal arithmetic problems. Although Beilock et al.'s participants worked horizontal modular arithmetic problems, the justification for their account (Trbovich & LeFevre, 2003) and the account itself focus exclusively on the subtraction component of these problems. Thus, in Experiment 1, we identified a potential prepotent response for performance on horizontal subtraction problems, the method of adjustment. In Experiment 2, we showed that ST potentiated the use of this response. In Experiment 3, we pitted the mere effort account against Beilock et al.'s working memory account, and the results favored the mere effort account. Finally, in Experiment 4, we showed that if we controlled for the effect of the potentiated prepotent response, instead of performing more poorly, threatened participants performed better than no threat participants on horizontal modular arithmetic problems, supporting mere effort, but not the working memory account.

2. Experiment 1

To identify the prepotent response, we examined how participants approach the subtraction component of Beilock et al.'s (2007) modular arithmetic problems when they are presented vertically and horizontally without ST. Trbovich and LeFevre (2003) argue that when problems are presented in a horizontal format, participants solve the problems by making adjustments that produce intermediate values (e.g., participants adjust the second number to the nearest 10, subtract the two numbers, and then add the adjustment), which involve verbal codes that require the phonological loop, an approach that we will term the method of adjustment. Trbovich and LeFevre argue that participants use the visuospatial sketchpad to work the vertical problems, approaching them in the "traditional" manner as though they are working them on paper, starting on the right with the unit's place and then moving to the decade's (i.e., units to decades).

In the present experiment, female participants were presented with subtraction problems in either a horizontal or a vertical format. For both horizontal and vertical problems, participants were told that they could enter the answers starting with the left digit (i.e., from left-to-right) or starting with the right digit (i.e., from right-to-left).

If participants solve horizontal subtraction problems using the method of adjustment (Trbovich & LeFevre, 2003), we should find that the answers to these problems are entered from left-to-right because using the method of adjustment produces the entire answer. For example, to solve the problem $57 - 28$ using the method of adjustment, one rounds 28 to 30, then subtracts 30 from 57, and adds 2 to 27, yielding the whole answer, 29. In contrast, if one uses the visuospatial sketchpad to solve vertical problems, the "traditional" method would be used and answers would be entered from right-to-left. That is, to solve $57 - 28$, one would borrow a ten from the decades column to perform the subtraction $17 - 8$, which equals 9. One could then enter the 9 in the right hand box, followed by the decade operation, $4 - 2 = 2$, which would be entered in the left hand box. We tested this direction of digit entry hypothesis in Experiment 1. If supported, it suggests the approach that participants take to horizontal problems, which would then provide a candidate for the prepotent response on this problem type.

The primary dependent measure in this experiment was direction-of-entry of the digits. In terms of actual performance, participants self-select the direction in which they enter the digits, which could affect their performance.

2.1. Method

2.1.1. Participants

Twenty-nine females[3] from a northeastern university participated in the study for course credit. In this and the subsequent experiments, we did not collect any demographic information other than gender.

2.1.2. Procedure

Participants came into the lab individually and were told that they would be solving math problems. The task consisted of 160 high-demand subtraction problems. As described by Beilock et al. (2007), these high-demand problems were two-digit subtraction problems that involved numbers greater than 10 and a borrow operation. Problems were presented either horizontally or vertically.

Participants completed the problems on a computer. For the participants who saw horizontal problems, two boxes were presented to the right of the problem (e.g., $57 - 28 = \square\,\square$). For participants who saw vertical problems, the boxes were beneath the numbers. Problems were presented in random order and remained until the participant answered after which there was a 750 ms intertrial interval. Participants were told that the program would tell them when they had finished the task. In actuality, the program always stopped participants after 10 min.

All participants were told that they could enter the answer from right-to-left or from left-to-right. To indicate what digit they wanted to enter first, they just needed to click on that box first (see Supplemental materials for full instructions). Before starting, participants completed 5 practice problems.

Upon completion of the math task, participants were asked about the approach(es) that they took to solving the problems. They read: "In working on the problems (e.g., $57 - 38$), you may have followed a traditional approach (e.g., started on the right; to subtract the 8 from the 7, you borrowed one from the 5 in the tens column making the 7 into 17; $17 - 8 = 9$; subtracted the 3 from the 4 in the tens column yielding a 1; and then put the number together, 19). Or you could have used some other approach. For example, you could have added 2 to 38 (40), subtracted 40 from 57, and then added back the 2 to get the final answer, 19. Or you could have used some variant of this approach. On what percentage of the problems did you use the traditional approach? ______% On what percentage of the problems did you use some other approach? _______% The two percentages should sum to 100%. If you used a non-traditional approach, please describe the approach or approaches that you used."

[3] We ran an initial study using modular arithmetic problems to attempt to replicate Beilock et al.'s (2007) results in a paradigm that was sensitive to potential motivational effects (i.e., participants were given more problems than could be solved in the time allotted). Just as Beilock et al., we found that threatened females performed more poorly than control females, and also found that males were not affected by the threat manipulation. Given this finding, in subsequent research we used only female participants.

2.2. Results

All of the results were analyzed in one-way ANOVAs with orientation (horizontal vs. vertical) as the independent variable. Beilock et al. (2007) excluded data from analyses if participants solved less than 75% of the problems correctly. However, all participants met this 75% criterion.

2.2.1. Task performance

An analysis of the proportion of digits entered from left-to-right by each participant yielded a significant effect, $F(1, 27) = 4.15$, $p = .05$, $\eta_p^2 = .25$. The proportion of horizontal problems entered by each participant from left-to-right ($M = .62$, $SD = .48$) was greater than the proportion of vertical problems ($M = .28$, $SD = .43$). Of course, the right-to-left analysis is a mirror image of left-to-right. Thus, participants tended to enter horizontal problems from left-to-right and vertical problems from right-to-left.

For math performance, there was no reliable effect for percentage correct (Overall $M = 94\%$). The analysis of number of problems solved correctly yielded a significant effect for orientation, $F(1, 27) = 5.52$, $p < .05$, $\eta_p^2 = .17$. Participants solved more problems correctly when they were presented in a vertical orientation ($M = 78.71$, $SD = 24.14$) than when they were presented in a horizontal orientation ($M = 61.53$, $SD = 14.33$).[4]

2.2.2. Solution approach measure

This analysis yielded an effect for orientation, $F(1, 27) = 4.57$, $p < .05$, $\eta_p^2 = .14$. Participants reported using the traditional method on a smaller proportion of horizontal problems ($M = .31$, $SD = .44$) than participants working on vertical problems ($M = .66$, $SD = .45$). In other words, when working on horizontal problems, participants reported using the method of adjustment on 69% of the problems, whereas when working on vertical problems, they reported using the traditional method on 66% of them.

2.3. Discussion

Consistent with Trbovich and LeFevre's (2003) argument, on horizontal problems, the participants tended to enter the numbers from left-to-right and reported that they used the method of adjustment, suggesting that they calculated the entire answer and then entered it. Or they reported some variant of this approach in which they adjusted one of the numbers up or down, and then performed the subtraction, followed by taking into account the adjustment.[5] This finding is consistent with Trbovich and LeFevre's proposal that the participants' solution approach on horizontal problems makes use of the phonological loop to store intermediate values, as suggested by Beilock et al. (2007). On vertical problems, we found that participants tended to enter the answers from right-to-left, and reported using the traditional method (i.e., the unit-to-decade algorithm that is used on paper).

These findings suggest that the method of adjustment is the approach that most participants take to solving the horizontal subtraction problems. Given that this is the prepotent response, the mere effort account would argue that this tendency would be potentiated when female participants are subject to threat. We examined this hypothesis in the next experiment.

3. Experiment 2

To test the potentiation hypothesis, we added a ST manipulation to the paradigm used in Experiment 1. Thus, females were either threatened or not and then solved horizontal or vertical subtraction problems for which they could enter the answer either from left-to-right or right-to-left.

If using the method of adjustment is the prepotent response for horizontal subtraction problems, we should find that threatened females enter the answers from left-to-right more than unthreatened participants. Using the method of adjustment yields the whole answer, which is most easily entered from left-to-right. In contrast, when one uses the traditional method, the problem is worked from right-to-left, and the most efficient mode of entry would be from right-to-left.

In this experiment, as in Experiment 1, we primarily focused on the approach the participants took to solving the problems, rather than on performance. However, it would be consistent with the mere effort account to find that threatened participants solve fewer problems than control participants because the control participants tend to enter the answers from right-to-left more than the threatened participants. In this paradigm, entering the numbers from right-to-left confers an advantage because, once the right hand number is entered, there is no need to remember it.

The working memory account does not make direction-of-entry predictions. However, it is easier to enter the numbers from right-to-left, and if threatened participants did so, it would reduce or eliminate the debilitation in that they would not need the phonological loop to keep track of that intermediate value.

3.1. Method

3.1.1. Participants

Seventy-four female undergraduates from a northeastern university participated in the study for course credit.

3.1.2. Procedure

The procedure and measures were the same as those used in Experiment 1 with the addition of a ST manipulation. ST participants were told, “the task on which you are about to participate has been shown to produce gender differences” (see Supplemental materials for full manipulations). No ST participants were told, “The task on which you are about to participate has not been shown to produce gender differences.” Previous research using a similar manipulation has shown that it produces ST effects (e.g., Brown & Pinel, 2003).

Following task completion, participants filled out the following questions as a manipulation check: To what extent are there gender differences in performance on this task? (1 = *no gender differences* and 11 = *gender differences*); and Who do you believe performs better on this task? (1 = *males perform better*, 6 = *males and females perform the same*, and 11 = *females perform better*). In addition, they responded to the same questions about the approaches they took to solving the problems as those used in Experiment 1.

3.2. Results

All of the results were analyzed in two-way ANOVAs with threat (ST vs. no ST) and orientation (horizontal vs. vertical) as independent variables. Two participants failed to meet Beilock et al.'s (2007) 75% accuracy criterion (1 participant in the ST/horizontal condition and 1

[4] Past research on stereotype threat has used number of problems correct (e.g., Ben-Zeev et al., 2005), percentage correct (e.g., Beilock et al., 2007) or both (e.g., Steele & Aronson, 1995) to demonstrate threat effects. We instructed participants to be as accurate as possible while solving as many problems as they could. Because these problems were quite simple, participants achieved high levels of accuracy (from 93% to 97%) with no between condition differences. Differences were produced in number of problems correct, which were mirrored in problems attempted. Because “problems solved correctly” matches the goal given to our participants we used this measure as the primary dependent variable. Given the participants' instructions, if the participants in one condition correctly solved more problems than participants in another condition with no cost in accuracy, we argue that this clearly represents better performance.

[5] Although participants sometimes described other methods (e.g., go up by tens from the lower number until you get within units of the upper number and then adjust up or down), the great majority of the participants described using the traditional method and/or the method of adjustment. As a result, we focused on these methods.

participant in the no ST/horizontal condition), and were excluded from the analysis.

3.2.1. Manipulation checks

Participants in the threat condition reported that gender differences existed to a greater extent ($M = 5.85$, $SD = 1.93$) than no ST participants ($M = 2.82$, $SD = 2.19$), $F(1, 68) = 39.13$, $p < .001$, $\eta_p^2 = .37$. ST participants also reported that males perform better on the task than females to a greater extent ($M = 4.39$, $SD = 1.73$) than no ST participants ($M = 6.03$, $SD = 1.13$), $F(1, 68) = 21.50$, $p < .001$, $\eta_p^2 = .24$. These results indicated that the ST manipulation was successful.

3.2.2. Task performance

Analysis of the proportion of digits that each participant entered from left-to-right produced a Threat × Orientation interaction, $F(1, 68) = 3.88$, $p = .05$, $\eta_p^2 = .05$. To test the hypothesis that threat would potentiate the tendency to enter digits from left-to-right (use the method of adjustment), we contrasted the proportion of horizontal problems that each participant entered from left-to-right for threat and no threat participants, $F(1, 68) = 7.20$, $p < .05$, $d = .65$. Threat participants entered a greater proportion of problems from left-to-right ($M = .87$, $SD = .32$) than no threat participants ($M = .51$, $SD = .50$). The same contrast for vertical problems was not reliable, $p > .40$ ($M_{Threat} = .28$, $SD = .43$; $M_{No\ Threat} = .31$, $SD = .46$). This analysis also revealed a main effect for orientation, $F(1, 68) = 15.57$, $p < .001$, $\eta_p^2 = .19$. As in Experiment 1, the proportion of horizontal problems entered by each participant from left-to-right ($M = .71$, $SD = .44$) was greater than the proportion of vertical problems ($M = .29$, $SD = .44$).

The analysis for math performance revealed no reliable differences for percentage correct (Overall $M = 96\%$). The analysis of number of problems solved correctly yielded a Threat × Orientation interaction, $F(1, 68) = 9.63$, $p < .01$, $\eta_p^2 = .12$. Once again, planned contrasts (Kirk, 1995) were used to compare threat conditions within this two-way analysis.

Threatened participants solved fewer horizontal problems ($M = 64.46$, $SD = 15.54$) than non-threatened participants ($M = 80.41$, $SD = 30.20$), $F(1, 68) = 4.92$, $p < .05$, $d = .54$. However, threatened participants solved more vertical problems ($M = 99.65$, $SD = 26.07$) than participants who were not threatened ($M = 82.75$, $SD = 14.69$), $F(1, 68) = 4.74$, $p < .05$, $d = .53$. There was also a main effect for orientation, $F(1, 68) = 12.57$, $p < .01$, $\eta_p^2 = .16$, as a result of the fact that participants correctly solved more vertical problems ($M = 91.46$, $SD = 22.68$) than horizontal problems ($M = 71.41$, $SD = 24.12$).

3.2.3. Solution approach measure

Analysis of the percentage of problems on which the participants self-reported using the traditional method revealed a significant Threat × Orientation interaction, $F(1, 68) = 3.89$, $p = .05$, $\eta_p^2 = .05$. A planned contrast (Kirk, 1995) showed that threat participants reported using the traditional method on a smaller proportion of horizontal problems ($M = .18$, $SD = .32$) than no threat participants ($M = .51$, $SD = .45$), $F(1, 68) = 6.86$, $p < .05$, $d = .64$. The same contrast for vertical problems was not reliable, $p > .40$, ($M_{Threat} = .74$, $SD = .40$; $M_{No\ Threat} = .71$, $SD = .42$). This analysis also revealed a main effect for orientation, $F(1, 68) = 16.64$, $p < .001$, $\eta_p^2 = .20$. As in Experiment 1, participants who solved horizontal problems reported using the traditional method less ($M = .32$, $SD = .41$) than participants who worked on vertical problems ($M = .72$, $SD = .40$).

3.3. Discussion

Replicating Experiment 1, and consistent with the argument that using the method of adjustment is the prepotent approach to solving horizontal subtraction problems, we found that participants tended to work from left-to-right using the method of adjustment on horizontal problems, but to work from right-to-left using the traditional method on vertical problems.

Consistent with the mere effort account, we found that the tendency to use the method of adjustment was potentiated for participants under threat. We also found that threatened participants solved fewer horizontal subtraction problems correctly than no threat controls. However, once again, the focus in this experiment was on direction-of-entry, not on performance. In this paradigm, there is a built-in advantage when one uses the traditional method and enters the numbers from right-to-left: The unit's digit, once calculated, need not be remembered; it can simply be entered in the unit's place. Because the no threat females used the traditional method and entered the digits from right-to-left more than the threatened females, this shortcut may account for their performance advantage.

On vertical problems, threatened females performed better than the no threat controls. However, on these problems, threatened and non-threatened females did not differ in the extent to which they reported using the traditional method (.74 and .71, respectively). Thus, there is no evidence that it was the potentiation of the prepotent response that produced the performance difference. It is possible that the traditional approach to solving vertical problems is so well learned that the prepotent response is at a ceiling. However, even if threatened and non-threatened participants used the traditional method to the same extent, the mere effort account would predict that the greater motivation of the threatened females would lead them to outperform the controls, and this is what was found. This finding is consistent with the findings in Jamieson and Harkins's (2007) prosaccade condition in which threatened and non-threatened females looked at the cue to the same extent (prepotent response at a ceiling), but threatened females identified the orientation of the target more quickly than non-threatened ones.

In any event, Beilock et al.'s (2007) argument is that threat debilitates performance on the horizontal modular problems because solving the subtraction portion requires phonological resources (e.g., remembering intermediate values) that are taken up by the thoughts and worries produced by threat. Thus, in Experiment 3, we directly pitted the mere effort account against this working memory account on these problems.

4. Experiment 3

Females were randomly assigned to either a ST or no threat condition. Crossed with this manipulation, participants were asked to solve horizontal or vertical two digit subtraction problems. Crossed with these two manipulations, participants were required to enter the answers to these problems either from right-to-left or from left-to-right.

In this paradigm, Beilock et al.'s (2007) working memory account would predict that, on horizontal problems, requiring participants to enter their answers from right-to-left could reduce the debilitation produced by ST. Participants do not need to use the phonological loop for remembering the intermediate value (Imbo, Vandierendonck, & Vergauwe, 2007). They can simply perform the unit's subtraction and enter the value in the right hand box. As a result, the fact that the phonological loop is occupied by concerns about fulfilling the threat should have diminished or no impact on performance (see Fig. 1). However, this shortcut is not available when answers are entered from left-to-right. Because the whole answer is entered, the phonological loop is needed to keep track of the intermediate values. According to Beilock et al.'s account, thoughts about fulfilling the stereotype should interfere with remembering these values, resulting in debilitated performance (see Fig. 1).

In contrast, the mere effort account would predict that ST would potentiate the prepotent response, the method of adjustment. Entering the answer from left-to-right is consistent with the method of adjustment, which produces the whole answer. Thus, the fact that the threat participants are motivated to want to perform well should lead them to outperform no threat participants on horizontal problems when

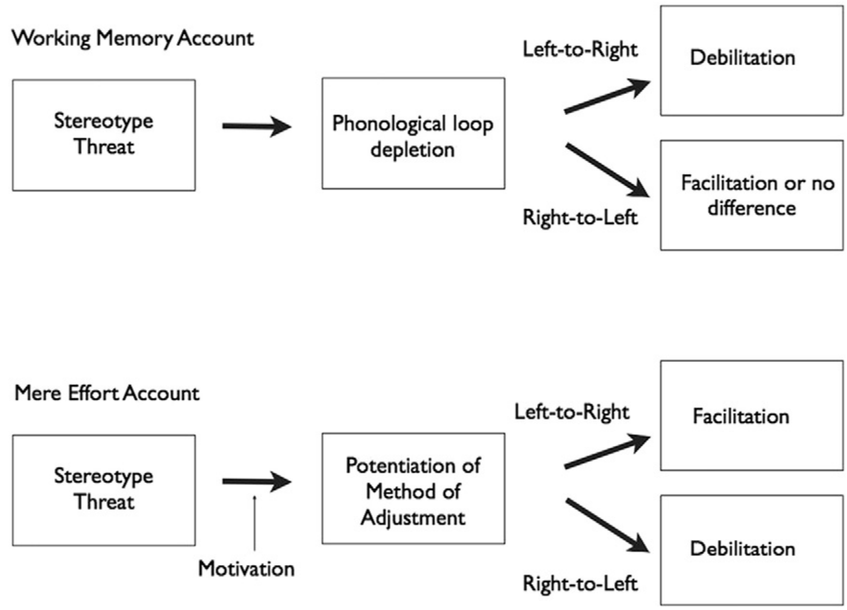

Fig. 1. Experiment 3: Predictions for the working memory and mere effort accounts for horizontal subtraction performance when participants were forced to enter their answers from left-to-right or right-to-left.

they enter the answers from left-to-right (see Fig. 1). On the other hand, using the method of adjustment should debilitate performance when the answer is entered from right-to-left. One must remember the adjustments to reach the final answer and then enter this answer from right-to-left, even though it would usually be written from left-to-right. Also, the fact that no threat participants are more likely than threat participants to use the traditional method confers the advantage described above when the digits must be entered from right-to-left (i.e., they do not need to remember the intermediate value and can enter it into the unit's box, making problem solution easier). Thus, the potentiated use of the method of adjustment should lead the threatened participants to perform more poorly than no threat participants when they enter the answers from right-to-left (see Fig. 1).

To the extent that participants rely more heavily on visuospatial resources to solve vertically presented problems, verbal working memory resources (i.e., the phonological loop) are not necessary (Beilock et al., 2007). As a result, the working memory account would not make a strong prediction for vertical problems in the current experimental paradigm. However, it would be consistent with a motivational account, like mere effort, to find that threatened participants perform better on vertical problems than no threat participants, as was found on the vertical subtraction problems in Experiment 2.

4.1. Method

4.1.1. Participants

One hundred and sixteen female undergraduates from a northeastern university participated in the study for course credit.

4.1.2. Procedure

The same 160 horizontal and vertical subtraction problems used in Experiment 2 were used in the current experiment. The procedures were identical to those used in Experiment 2 with one exception. Participants were instructed to enter their answers from left-to-right or from right-to-left. As part of the direction-of-entry manipulation, the computer cursor was programmed to always be on the appropriate box. For example, if participants were asked to enter their answer from right-to-left, then the cursor was always on the right hand box and the participant was told to enter that digit before it moved on to the left hand box (see Supplemental materials for full instructions). After the practice problems, the ST manipulation used in Experiment 2 was implemented.

After completing the task, participants responded to the same threat manipulation check items that were used in Experiment 2, and answered the same questions about the approaches they took to solving the problems as were used in Experiments 1 and 2.

4.2. Results

All of the results were analyzed in three-way ANOVAs with threat (ST vs. no ST), orientation (horizontal vs. vertical), and direction (left-to-right vs. right-to-left) as independent variables. Two participants failed to reach Beilock et al.'s (2007) 75% criterion (1 participant in the ST/horizontal/left-to-right condition and 1 participant in the no ST/horizontal/left-to-right condition) and were excluded from analyses.

4.2.1. Manipulation checks

Main effects for ST condition were found for both questions but no other reliable effects were found. Participants subject to ST reported that gender differences existed to a greater extent ($M = 6.37$, $SD = 2.44$) than no ST participants ($M = 2.43$, $SD = 1.98$), $F\,(1, 106) = 88.33$, $p < .001$, $\eta_p^2 = .46$. ST participants also reported that males perform better on the task than females to a greater extent ($M = 4.55$, $SD = 1.94$) than no ST participants ($M = 5.98$, $SD = .96$), $F\,(1, 106) = 22.20$, $p < .001$, $\eta_p^2 = .17$. These results indicated that the ST manipulation used was successful.

4.2.2. Task performance

There were no reliable effects for percentage correct (Overall $M = 96\%$). On problems correct, results revealed a three-way interaction, $F\,(1, 106) = 5.21$, $p < .05$, $\eta_p^2 = .05$. To decompose the three-way interaction, two two-way ANOVAs were conducted, one for horizontal problems and one for vertical problems, using the overall error term for the three-way interaction for each analysis. We took this approach only after we ensured that we had not violated the assumption of homoscedasticity of errors. This approach is the same as running two linear models that include the three-way interaction but use dummy codes to focus the two-way interaction of interest on only the horizontal or vertical condition.

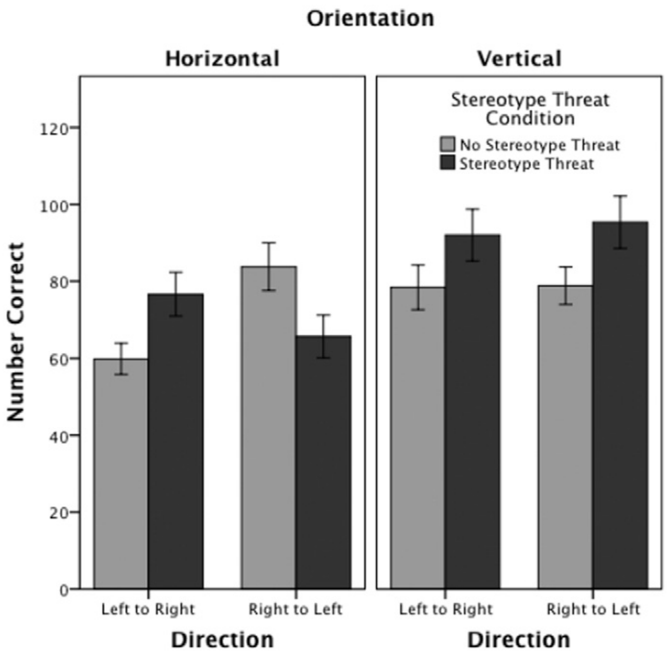

Fig. 2. Experiment 3: Mean number of problems correct on horizontal and vertical problems as a function of ST condition and direction of answer entry. Note: Error bars = ± standard error of the mean.

Analysis of the horizontal problems in a 2 (threat vs. no threat) × 2 (left-to-right vs. right-to-left) ANOVA revealed only a Threat × Direction interaction, F (1, 106) = 9.29, $p < .01$, $\eta_p^2 = .08$. Planned contrasts (Kirk, 1995) were used to compare threat conditions within this two-way analysis.

As seen in Fig. 2, threatened participants solved more horizontal problems correctly (M = 76.61, SD = 24.17) than non-threatened participants (M = 59.85, SD = 14.49) when the problems were entered from left-to-right, F (1, 106) = 4.38, $p < .05$, d = .41. However, non-threatened participants solved more problems correctly (M = 83.79, SD = 23.26) than threatened participants (M = 65.67, SD = 21.45) when they were entered from right-to-left, F (1, 106) = 4.91, $p < .05$, d = .43.

Analysis of the vertical problems in a 2 (threat vs. no threat) × 2 (left-to-right vs. right-to-left) ANOVA revealed only a main effect for threat, F (1, 106) = 6.30, $p < .05$, $\eta_p^2 = .06$. Participants subject to ST solved more vertical problems correctly (M = 93.74, SD = 24.51) than participants who were not subject to threat (M = 78.63, SD = 19.48) (see Fig. 2).

4.2.3. Solution approach measure

This analysis revealed three main effects. Threatened participants reported using the traditional method on a smaller proportion of problems (M = .34, SD = .44) than non-threatened participants (M = .60, SD = .43), F (1, 106) = 9.93, $p < .05$, $\eta_p^2 = .09$. Participants working on horizontal problems reported using the traditional method on a smaller proportion of problems (M = .34, SD = .44) than participants who worked on vertical problems (M = .60, SD = .46), F (1, 106) = 10.07, $p < .05$, $\eta_p^2 = .09$. And participants reported using the traditional method on a smaller proportion of problems entered from left-to-right (M = .36, SD = .42) than participants who entered their answers from right-to-left (M = .57, SD = .47), F (1, 106) = 6.47, $p < .05$, $\eta_p^2 = .06$.

4.3. Discussion

In Experiment 3, threatened and non-threatened participants were required to enter the answers to horizontal or vertical subtraction problems from left-to-right or right-to-left. On horizontal problems, we found that threatened participants performed better than controls when answers were entered from left-to-right, but worse when the answers were entered from right-to-left. This pattern of findings is consistent with mere effort, but not the working memory account.

Also consistent with a motivational account, and with the findings for vertical subtraction problems in Experiment 2, in Experiment 3 we found that threat participants outperformed no threat controls on vertical problems (cf. Jamieson & Harkins, 2007).

Although these findings support the mere effort account, but not the working memory explanation, for performance debilitation on horizontal subtraction problems, they do not explain why threatened participants perform more poorly than non-threatened participants on Beilock et al.'s (2007) horizontal *modular arithmetic* problems. In fact, the designs of both Experiments 2 and 3 incorporate features that permit tests of hypotheses that are relevant to performance on horizontal subtraction problems but are not present in horizontal modular arithmetic problems.

For example, in Experiment 3, we found that when participants were required to enter the problems from right-to-left, non-threatened participants outperformed threatened participants. For the non-threatened participants, the prepotent response was not potentiated, making it more likely that they would recognize that they need not remember the unit's digit; they could simply perform the unit's subtraction and enter the value in the right hand box. However, this shortcut is not available to them when they solve horizontal modular arithmetic problems. Instead, using the traditional, units-to-decade method working from right-to-left requires that they remember the unit's digit when they move to the decade calculation.

Even so, the traditional method may be more efficient than the method of adjustment. For example, when using the traditional method, the steps may be more easily routinized than when using the method of adjustment. One subtracts the unit's digits, borrowing a ten (e.g., 57 − 28; 17 − 8 = 9), and then decrements the decade minuend by one and subtracts the subtrahend (4 − 2 = 2), yielding the answer (29). To use this method effectively, one need only know the subtraction facts for numbers under 20, and then remember the unit's digit while decrementing the decade's value. In contrast, when using the method of adjustment, for each problem one must determine how much to increment the subtrahend (57 − 28; plus 2, 28 to 30), then subtract this value from the minuend (57 − 30 = 27), and then add the adjustment (27 + 2 = 29). The amount to adjust the subtrahend will change with each problem, as will the size of the difference between the minuend and the adjusted subtrahend. Finally, one must remember to make the adjustment, as well as remember what it is. Thus, it is possible that the traditional method is more efficient than the method of adjustment even without the advantage conferred by the opportunity to enter the unit's digit first. If this were the case, because the non-threatened participants use the traditional method more than threatened participants, it could account for why they perform better than threatened participants on horizontal modular arithmetic problems. This hypothesis was tested in Experiment 4.

5. Experiment 4

Females were randomly assigned to either a ST or no ST condition. Crossed with this manipulation, participants were asked to solve the horizontal subtraction portion of the problems under one of three instruction sets: They were asked to use the traditional, units-to-decades method; they were asked to use the method of adjustment; or they were given no instructions.

In the no instruction, replication condition, we should replicate Beilock et al.'s (2007) finding that non-threatened participants solve more modular arithmetic problems than threatened participants with no cost in accuracy. Our argument is that this difference stems from the fact that threatened participants use the potentiated prepotent response, the method of adjustment, more than control participants, and this method is less efficient than the traditional, right-to-left approach. If this is the case, when the solution method is held constant, the more motivated, threat participants should outperform the unthreatened participants, whether the participants are instructed to use the method of adjustment or the traditional method. That is, just as was the case for Jamieson and Harkins's (2007) participants on the prosaccade task, and the current participants on the vertical problems, when the approach is held constant, the more motivated threat participants will outperform the no threat participants. In addition we should find that, overall, more problems are solved correctly using the traditional method, reflecting the fact that this approach is more efficient.

5.1. Method

5.1.1. Participants

Eighty-seven female undergraduates from a northeastern university participated in the study for course credit.

5.1.2. Procedure

Participants completed modular arithmetic problems (e.g., 52 = 24 (mod 3)). They were to subtract the second number from the first, divide by the mod, and then report, true or false, whether the result was a whole number. Consistent with the methodology reported by Beilock et al. (2007), half of the modular arithmetic problems were "true" and each "true" problem had a "false" correlate that only differed as a function of the number involved in the division (mod) statement. For instance, if the "true" problem 51 = 19 (mod 4) was presented, then a "false" correlate problem 51 = 19 (mod 3) was also presented at some point in the problem set. This pairing was conducted to equate "true" and "false" problems as much as possible in terms of the specific numbers used in each equation.

Because all "false" problems were correlates of a "true" problem, we sought to reduce any effect familiarity had on performance. Problems were randomized throughout the test with the constraint that if a particular problem appeared on a page, then its correlate could not appear until the participant had completed at least 10 problems in between. One randomized version of the problems was made and given to all participants.

The paper and pencil problem sets consisted of 120 modular arithmetic problems, with 10 problems per page. Participants were asked to solve as many problems as they could, as accurately as possible, and were told that the problems were to be solved in their heads without using scratch paper. After the task description, participants were randomly given one of three instructions. In the method of adjustment condition, participants were asked to increase the subtrahend to the next ten, perform the subtraction, and then add the adjustment. In the traditional condition, participants were asked to subtract the unit's digits by borrowing a ten, and then to decrement the decade minuend by one and subtract the subtrahend. In each of these cases, the participants were walked through an example problem (see Supplemental materials for full instructions). In the replication condition, participants were given no instructions about how to perform the subtraction, and the experimenter simply solved the problem for them. After these instructions, participants were asked to complete 8 practice problems to familiarize themselves with the task. Once the practice problems were completed, the ST manipulation was induced.

After completing the 10-minute task, participants responded to threat manipulation check items and solution approach questions used in Experiments 2 and 3.

5.2. Results

All of the results were analyzed in two-way ANOVAs with threat (ST vs. no ST) and instruction (replication vs. method of adjustment vs. traditional method) as independent variables. Three participants failed to reach Beilock et al.'s (2007) 75% accuracy criterion (two participants in the ST/replication condition and one in the no ST/replication condition) and were excluded from analyses.

5.2.1. Manipulation checks

Responses to the questions asking to what extent gender differences exist on the task and which gender performs better on the task produced main effects for ST on both questions but no other reliable effects. Participants subject to ST reported that gender differences existed to a greater extent ($M = 6.27$, $SD = 2.03$) than no ST participants ($M = 2.21$, $SD = 1.61$), $F(1, 78) = 104.84$, $p < .001$, $\eta_p^2 = .57$. ST participants also reported that males perform better on the task than females to a greater extent ($M = 3.98$, $SD = 2.02$) than no ST participants ($M = 5.72$, $SD = 1.14$), $F(1, 78) = 23.60$, $p < .001$, $\eta_p^2 = .23$. These results indicated that the ST manipulation was successful.

5.2.2. Task performance

There were no reliable effects for percentage correct on the math problems (Overall $M = 96\%$). Analysis of the number of problems correct revealed a Threat × Instruction interaction, $F(2, 78) = 6.55$, $p < .01$, $\eta_p^2 = .14$. Planned contrasts (Kirk, 1995) were used to compare threat and no threat means within each instruction condition.

As seen in Fig. 3, in the no instruction condition no threat participants solved significantly more problems ($M = 65.67$, $SD = 20.10$) than threat participants ($M = 53.36$, $SD = 13.13$), $F(1, 78) = 4.12$, $p < .05$, $d = .46$. This is the same debilitation effect reported by Beilock et al. (2007). However, in the method of adjustment condition, threatened participants solved more problems correctly ($M = 58.71$, $SD = 13.36$) than no threat participants ($M = 42.93$, SD = 10.19), $F(1, 78) = 6.55$, $p < .05$, $d = .58$. Likewise, in the traditional method condition, threatened participants ($M = 81.92$, $SD = 19.15$) also outperformed no threat participants ($M = 68.21$, $SD = 19.24$), $F(1, 78) = 4.76$, $p < .05$, $d = .49$.

This analysis also revealed a main effect for instruction, $F(2, 78) = 15.52$, $p < .0001$, $\eta_p^2 = .29$. Consistent with the argument that the traditional approach is more efficient than the method of adjustment,

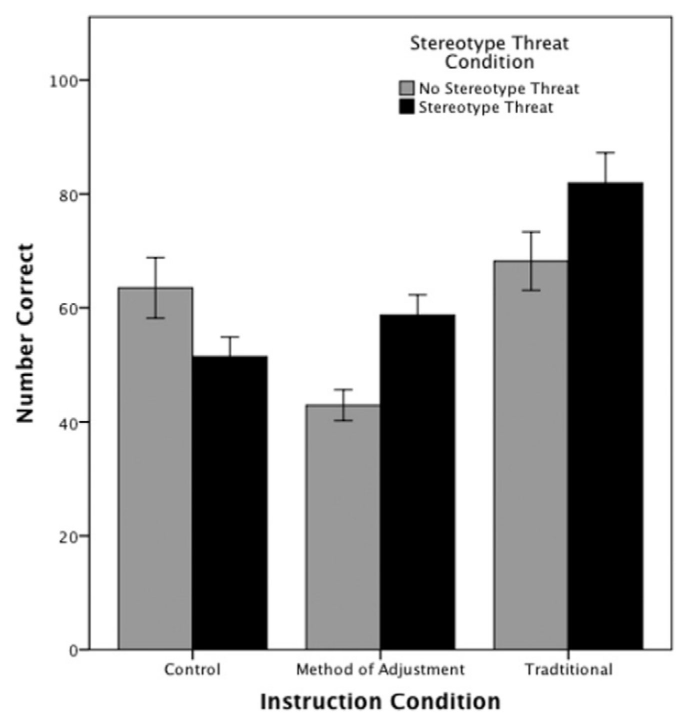

Fig. 3. Experiment 4: Mean number of problems correct on horizontal modular arithmetic problems as a function of ST and instruction condition. Note: Error bars = ±standard error of the mean.

participants using the traditional method solved more problems correctly ($M = 74.82$, $SD = 20.08$) than participants using the method of adjustment ($M = 50.82$, $SD = 14.16$), $F(1, 78) = 30.32$, $p < .0001$, $d = 1.25$. The mean in the no instruction, replication condition ($M = 59.72$, $SD = 17.92$) differed significantly from each of the instruction means (Method of adjustment: $F(1, 78) = 4.04$, $p < .05$, $d = .46$; Traditional method: $F(1, 78) = 12.69$, $p < .0001$, $d = .81$).

5.2.3. Solution approach measure

This analysis revealed a significant instruction effect, $F(1, 78) = 121.13$, $p < .0001$, $\eta_p^2 = .76$. Consistent with the success of the instruction manipulation, participants instructed to use the traditional method reported doing so significantly more ($M = .95$, $SD = .08$) than participants asked to use the method of adjustment ($M = .02$, $SD = .10$), $F(1, 78) = 234.27$, $p < .0001$, $d = 3.47$. The mean for participants in the no instruction condition ($M = .33$, $SD = .37$) differed significantly from each of the instruction means (Method of adjustment: $F(1, 78) = 26.99$, $p < .0001$, $d = 1.18$; Traditional method: $F(1, 78) = 105.83$, $p < .0001$, $d = 2.33$).

5.3. Discussion

When given no instructions about how to solve the problems, we replicated Beilock et al.'s (2007) findings: Participants in the no threat condition solved more problems correctly than participants in the threat condition with no cost in accuracy. However, when the participants were told which approach to use to solve the problems, this difference was reversed. Whether told to use the method of adjustment or the traditional method, threatened participants solved more problems correctly than no threat participants.

We argue that when given no instructions, threat participants did not perform as well as no threat participants because the prepotent response, using the method of adjustment, was potentiated by ST, leading these participants to use this method more than no threat participants. And, as shown in Experiment 4, this method is less efficient than the traditional, units-to-decades method. When this difference in the use of the methods is eliminated through instruction, the performance difference is not only eliminated, but reversed, reflecting the greater motivation of the threat participants.

6. General discussion

In 1995, Steele and Aronson coined the term “stereotype threat” to describe the concern that arises when one feels at risk of confirming a negative stereotype about one's group. In the 20 subsequent years, the overwhelming majority of research has found that this concern debilitates performance on cognitive tasks. The currently prevailing explanation for this finding is the working memory interference account (e.g., Schmader et al., 2008, 2015). According to this account, concern over confirming the stereotype produces thoughts and worries that utilize cognitive resources that could be devoted to task performance and it is this reduction in working memory capacity that produces the performance debilitation.

Using Baddeley's (1986, 2000) multicomponent model of working memory, Beilock et al. (2007) found support for this account by showing how ST undermines performance on horizontal modular arithmetic problems. Beilock et al. argued that the thoughts and worries produced by ST occupy the phonological loop, which is also required to store the intermediate values that are produced when participants solve the horizontal subtraction portion of these problems (Trbovich & LeFevre, 2003). It is this reduced capacity that debilitates performance.

In contrast, the mere effort account argues that ST motivates stigmatized individuals to disconfirm the stereotype, which potentiates the prepotent response on the given task (Jamieson & Harkins, 2007, 2009; Seitchik et al., 2014). If the prepotent response is correct, ST facilitates performance for stigmatized individuals. 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 have the knowledge that their prepotent tendencies are incorrect and are given the opportunity to correct them, performance can be facilitated (Jamieson & Harkins, 2007, 2009).

Beilock et al.'s (2007) explanation focuses on performance of the subtraction component of the modular arithmetic problems. In Experiment 1, we found that participants tended to enter answers to the horizontal subtraction problems from left-to-right and reported using the method of adjustment. On vertical problems, participants tended to enter the answers from right-to-left and reported using the traditional method. In Experiment 2, we found that the tendency to use the method of adjustment on horizontal subtraction problems, the prepotent response, was potentiated by ST: Threatened participants entered a greater proportion of their answers from left-to-right than non-threatened participants. In addition, participants under ST reported using the method of adjustment on horizontal problems more than control participants.

Experiment 3 pitted mere effort against the working memory account by requiring participants to enter the answers from left-to-right or from right-to-left. Consistent with mere effort, but not the working memory deficit account, we found that threatened participants performed better than control participants when digit entry was from left-to-right, but worse than control participants when digit entry was from right-to-left.

In Experiment 4, we used modular arithmetic to test the hypothesis that threatened females performed more poorly than no threat females on these problems because the former participants used the method of adjustment on the subtraction portion of the problems more than the latter, and this method is less efficient than the traditional method. Consistent with this argument, we found that, overall, participants instructed to use the traditional method solved more problems correctly than participants instructed to use the method of adjustment. In accord with the motivational account, we also found that when solution approach was held constant, threatened females outperformed no threat females whether they used the method of adjustment or the traditional method.

Thus, these experiments support a motivational, mere effort, explanation for performance on horizontal modular arithmetic over Beilock et al.'s (2007) working memory interference account. However, Beilock and DeCaro (2007) suggest another way that a working memory deficit could account for performance debilitation on modular arithmetic problems. Because worries occupy working memory, threatened participants may try to use mathematical shortcuts to solve these problems, rather than trying to use their limited capacity to store intermediate values. For example, threatened participants may note that the response is likely to be true when all of the numbers in the problem are even (e.g., $(48 = 24) \bmod 6$), a shortcut that would use minimal working memory. Perhaps the use of the method of adjustment could represent such a shortcut, which happens to be ineffective.

However, the method of adjustment explicitly uses working memory. That is, using this method produces intermediate values that must be remembered, and so, unlike the “even number” shortcut described above, it does not avoid the use of working memory. In addition, in Experiment 4, the instructions to use the method of adjustment or the traditional method explicitly required participants to use working memory. We found that when solution approach was held constant, threat not only did not debilitate performance, but improved it. This finding is inconsistent with an account that argues that working memory is occupied by worries about fulfilling the stereotype, but it is consistent with the motivational account.

We also found evidence consistent with the motivational account on vertical problems. In this case, we did not find that threatened participants used the prepotent response, the traditional method, more than controls. For example, in Experiment 2, threat and no threat participants reported using the traditional method to the same extent ($M_{Threat} = .74$; $M_{No\ Threat} = .71$). Thus, there appears to be a very strong tendency for all participants to use the traditional method on vertical problems. However, even if participants used this approach to the same extent, we would expect to find that the greater motivation of threatened participants would lead them to outperform the control participants, and this is what was found (cf. performance on Jamieson and Harkins's (2007) prosaccade task). In Experiment 3, we found that threatened females outperformed control females on vertical subtraction problems, regardless of the direction-of-entry. This finding could simply reflect the fact that using the traditional method in the vertical orientation is so well learned that direction-of-entry does not matter. Of course, there are also other possibilities. For example, on vertical problems participants could use visual codes on the visuospatial sketchpad. Thus, although the findings for vertical problems are consistent with a motivational account, additional research will be required to determine exactly what process produces this effect.

In any event, the central claim of Beilock et al.'s (2007) account is that worries about confirming the stereotype occupy the phonological loop, and it is this loss of resources that produces debilitation on horizontal subtraction problems. The present research provides no support for this explanation, but does support the mere effort account both on horizontal subtraction problems and on horizontal modular arithmetic problems. Of course, one could argue Beilock et al.'s account represents only one version of a working memory account. However, any working memory account must argue that worries stemming from ST consume cognitive resources (e.g., Schmader et al., 2008, 2015), and in Experiment 4, we found that when participants were instructed to work the problems in ways that ensured that intermediate values would be produced, threatened females outperformed the controls. This finding, which is consistent with the mere effort account, is inconsistent with working memory accounts.

Other research has been cited (e.g., Mrazek et al., 2011; Schmader & Johns, 2003) as providing direct support for the working memory account (e.g., Schmader et al., 2015). However, key experiments in these papers use an intervening task methodology that Jamieson and Harkins (2011) have shown to be flawed. For example, in Experiment 1, Mrazek et al. told participants that they would be solving math problems and then implemented the ST manipulation. However, before the participants started the problems, they were asked to perform another task, performance on which was actually the primary dependent variable in the experiment. The researchers hypothesized that the negative thoughts produced by the threatened participants in anticipation of taking the math test would impact performance on this unrelated intervening task. As Mrazek et al. write: “A notable aspect of experiment 1 was that 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 one's group is negatively stereotyped” (p. 1245). However, Jamieson and Harkins (2011) have shown that participants perform poorly on the intervening task, not because negative thoughts produce working memory interference, but because they withdraw effort to save themselves for performance on the stereotyped task. In fact, Jamieson and Harkins (2011) have shown that when the stereotype applies to performance on the intervening task, consistent with mere effort, performance on the task (the Stroop in this case) is facilitated, not debilitated. A key experiment in Schmader and Johns's (2003, Experiment 3) work is subject to the same criticism.

Mere effort not only provides an account for the effect of ST on cognitive tasks but also on sensorimotor tasks. For example, Huber, Seitchik, Brown, Sternad, and Harkins (2015) asked participants to hit a virtual ball to a virtual line, using an actual racket, over the course of a series of 40-second trials. Previous research (e.g., Wei, Dijkstra, & Sternad, 2007) has shown that, on this task, novices have a tendency to hit the ball with positive acceleration, but over the course of the trials, they learn to hit the ball with negative acceleration, which allows them to minimize absolute error. Thus, the prepotent response on this task, hitting the ball with positive acceleration, is incorrect. Consistent with the mere effort account, Huber et al. found that female novices under ST hit the ball with positive acceleration to a greater extent than unthreatened females, resulting in greater absolute error. In a second experiment, females were given 12 blocks of practice on the task, after which the ST manipulation was introduced. By the 12th block, these participants had learned to hit with negative acceleration; that is, the prepotent response was correct. Consistent with the mere effort account, ST females performed better than control females on the measure of absolute error, as a result of the fact that they hit the ball with greater negative acceleration.

We argue that the present research in combination with previous work (e.g., Huber et al., 2015; Jamieson & Harkins, 2007, 2009) is consistent with the argument that mere effort can provide an account of ST performance effects, using the same core, motivational process to predict facilitation or debilitation on both cognitive and sensorimotor tasks. However, there are a number of issues that remain to be addressed in this account. For example, Schmader et al. (2015) suggest that mere effort and working memory deficit could work together to produce performance effects. They propose that it is when working memory is occupied by concerns about performance that participants revert to the prepotent response. Thus, the deficit in working memory leads the threatened participants to produce the prepotent response.

We would make two points about this possibility. First, it does not take into account the meaning of the term prepotent response. The prepotent response is the response that participants are most likely to make when they perform the given task. In research testing the mere effort account, this prepotency is established prior to the introduction of the threat manipulation. Thus, when threat is introduced, participants are not reverting to this response; it is the most likely response in the first place. Second, this explanation cannot account for performance on the horizontal modular arithmetic task used in the current research because working memory is needed to make the prepotent response. That is, the method of adjustment that is used on horizontal subtraction problems requires participants to use working memory to remember intermediate values.

Although this “two process” explanation does not appear to be capable of accounting for performance in the current research, it is possible that it could do so on other tasks. In addition, there are issues with the mere effort account itself that must be addressed. For example, the current line of research has focused exclusively on stereotypes concerning task performance by females, but past studies have shown (e.g., Steele & Aronson, 1995) that other stereotypes produce debilitation in performance. There is no reason to believe that the mere effort account would not hold for these other stereotypes as well, but this contention has yet to be tested.

Another issue stems from the fact that making mere effort predictions requires the identification of the prepotent response on the task under consideration. This requirement is a strength of this approach in that it appropriately emphasizes the importance of understanding the mechanics of the tasks that we use and also ties the account to specific predictions. On some tasks, the identification of the prepotent response is quite straightforward, as on the antisaccade task on which a specific response is produced (i.e., a reflexive saccade to the cue). In other cases, as in the current research, the prepotent response, the method of adjustment, involves a specific set of operations (e.g., participant adjusts the second number to the nearest 10, subtracts the two numbers, and then adds the adjustment), the contents of which differ from problem to problem.

In yet other cases, as in solving GRE quantitative problems (Jamieson & Harkins, 2009), the prepotent response may be more aptly described as an approach, rather than a specific response. On this task, based on previous research (Gallagher & De Lisi, 1994; Gallagher et al., 2000;

Quinn & Spencer, 2001), Jamieson and Harkins (2009) proposed that the participants' prepotent response was to try to use a formula or algorithm to solve the GRE problems, rather than logic and intuition.

It is true that this approach is consistent with the training that American students receive in mathematics (e.g., Katz, Bennett, & Berger, 2000), and thus represents the most likely response to a given problem. Nonetheless, this response is more molar than the response identified as prepotent on the horizontal subtraction problems, which is more molar than the response identified on the antisaccade task. This range in the type of responses identified as prepotent raises questions about the level at which we should expect to find the prepotent response on a given task that will need to be resolved in future research. Thus far, we have simply identified prepotent responses on a task-by-task basis, showing that each prepotent response is directly linked to the performance outcome on that task.

Even when the prepotent response for a given task has been identified and the findings for a given population of participants offer support for the mere effort account, caution must be exercised in generalizing the findings. For example, American students are typically not asked to memorize subtraction facts up to 100. If they were, they could rely on these subtraction facts to solve the horizontal subtraction problems instead of the method of adjustment, and this analysis would not apply.

Individual differences can also limit the scope of motivational explanations, such as the mere effort account. For example, in the domain of gender-math stereotypes, females must identify with their gender (Schmader, 2002) and with the domain (Smith & White, 2001) to be threatened in the context of the stereotype. Other research shows that participants under ST are quite willing to take advantage of explanations that will allow them to deflect responsibility for their performance (e.g., Ben-Zeev et al., 2005; Steele & Aronson, 1995; Stone, 2002). In these cases, participants may not perceive threat. Thus, the mere effort account applies to situations in which participants see the stereotype as directly relevant to them, and that do not provide a plausible explanation as to why they may not perform well. Under these circumstances, we argue that they are highly motivated to disconfirm the stereotype.

The present work along with previous research (e.g., Jamieson & Harkins, 2007, 2009; Kray, Thompson, & Galinsky, 2001; Seitchik et al., 2014) highlights the fact that females subject to threat do not simply accept their fate and resign themselves to performing poorly. They try very hard to disprove negative stereotypes directed at their group. Thus, the motivational processes argued for in this research offer a source of encouragement because the findings suggest that females react against being negatively typecast as poor performers in math and science.

Determining the mechanism(s) responsible for ST's influence on performance has practical implications. More specifically, ST in the domain of mathematics has been shown to reduce interest in science, technology, engineering, and mathematics (STEM) careers (e.g., Shapiro & Williams, 2012; Shapiro, Williams, & Hambarchyan, 2013; Smeding, 2012). This reduction in performance and interest may contribute to the gender gap in STEM-related majors and careers. In order to reduce this gender difference, it is important to understand *how* ST influences performance and create interventions to reduce this performance deficit. These interventions may then assist in reducing the gender gap in STEM-related fields. Mere effort makes a distinct contribution to this effort by showing that simple instructions can be enough to eliminate debilitation effects. For example, past research (e.g., Jamieson & Harkins, 2009) has shown that instructions about how to approach a problem can eliminate ST effects on the GRE math test. The same is true of the current research. 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 suggest otherwise, offering new approaches to improve performance in stigmatized individuals. Improved performance may then serve as positive reinforcement in STEM courses and increase interest in STEM careers. Thus, future research should not only focus on tests of the generalizability of the mere effort account but also on motivational interventions to overcome ST performance effects.

Acknowledgments

This research was funded by the U.S. Army Research Institute for the Behavioral and Social Sciences (Contract W5J9CQ-12-C-0046; PI: Stephen G. Harkins). 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2015.06.006.

References

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, *4*, 417–423. http://dx.doi.org/10.1016/S1364-6613(00)01538-2.

Beilock, S. (2010). *Choke: What the secrets of the brain reveal about getting it right when you have to.* New York: Simon & Schuster, Inc.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983–998. http://dx.doi.org/10.1037/0278-7393.33.6.983.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276. http://dx.doi.org/10.1037/0096-3445.136.2.256.

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.

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.

Cottrell, N. B. (1968). Performance in the presence of other human beings: Mere presence, audience, and affiliation effects. In E. C. Simmell, R. A. Hoppe, & G. A. Milton (Eds.), *Social facilitation and imitative behavior* (pp. 91–110). Boston: Allyn & Bacon.

Cottrell, N. B. (1972). Social facilitation. In C. G. McClintock (Ed.), *Experimental social psychology* (pp. 185–236). New York: Holt, Rinehart & Winston.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in Scholastic Aptitude Test: Mathematics problem solving among high-ability students. *Journal of Educational Psychology*, *86*, 204–211. http://dx.doi.org/10.1037/0022-0663.86.2.204.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–190. http://dx.doi.org/10.1006/jecp.1999.2532.

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. (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.

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

Imbo, I., Vandierendonck, A., & Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychological Research*, *71*, 467–483. http://dx.doi.org/10.1007/s00426-006-0044-8.

Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology*, *99*, 467–481. http://dx.doi.org/10.1037/a0018951.

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 effects 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 Bulletin*, *37*, 652–661. http://dx.doi.org/10.1177/0146167211399776.

Katz, I. R., Bennett, R. E., & Berger, A. E. (2000). Effects of response format on difficulty of SAT-mathematics items: It's not the strategy. *Journal of Educational Measurement*, *37*(1), 39–57. http://dx.doi.org/10.1111/j.1745-3984.2000.tb01075.x.

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

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. http://dx.doi.org/10.1037/0022-3514.80.6.942.

McFall, S., Jamieson, J. P., & Harkins, S. (2009). Testing the generalizability of 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.

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.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, *57*, 55–71. http://dx.doi.org/10.1111/0022-4537.00201.

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

Schmader, T., Hall, W., & Croft, A. (2015). Stereotype threat in intergroup relations. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson (Eds.), *APA handbook of personality and social psychology, Volume 2: Group processes* (pp. 447–471). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14342-017.

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*, 236–356. http://dx.doi.org/10.1037/0033-295X.115.2.336.

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

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, *66*, 175–183. http://dx.doi.org/10.1007/s11199-011-0051-0.

Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, *104*, 277–288. http://dx.doi.org/10.1037/a0030461.

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotype and stereotypes' connectedness to math performance. *Sex Roles*, *67*, 617–629. http://dx.doi.org/10.1007/s11199-012-0209-4.

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, *61*(6), 1040–1057. http://dx.doi.org/10.1177/00131640121971635.

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.

Trbovich, P. L., & LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory & Cognition*, *31*, 738–745. http://dx.doi.org/10.3758/BF03196112.

Wei, K., Dijkstra, T. M. H., & Sternad, D. (2007). Passive stability and active control in a rhythmic task. *Journal of Neurophysiology*, *98*, 2633–2646. http://dx.doi.org/10.1152/jn.00742.2007.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269–274. http://dx.doi.org/10.1126/science.149.3681.269.