

Girls can play ball: Stereotype threat reduces variability in a motor skill



Meghan E. Huber ^{a,*}, Adam J. Brown ^b, Dagmar Sternad ^{c,d,e}

^a Department of Bioengineering, Northeastern University, Boston, MA, USA

^b Department of Psychology, Northeastern University, Boston, MA, USA

^c Department of Biology, Northeastern University, Boston, MA, USA

^d Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA

^e Department of Physics, Northeastern University, Boston, MA, USA

ARTICLE INFO

Article history:

Received 28 October 2015

Received in revised form 1 April 2016

Accepted 17 May 2016

Available online xxxx

Keywords:

Stereotype threat

Mere effort

Motor control

Working memory

Motivation

ABSTRACT

The majority of research on stereotype threat shows what is expected: threat debilitates performance. However, facilitation is also possible, although seldom reported. This study investigated how stereotype threat influences novice females when performing the sensorimotor task of bouncing a ball to a target. We tested the predictions of two prevailing accounts for debilitation and facilitation due to stereotype threat effects: working memory and mere effort. Experimental results showed that variability in performance decreased more in stigmatized females than in control females, consistent with the prediction of the mere effort account, but inconsistent with the working memory account. These findings suggest that stereotype threat effects may be predicated upon the correctness of the dominant motor behavior, rather than on a novice-expert distinction or task difficulty. Further, a comprehensive understanding should incorporate the fact that stereotype threat can facilitate, as well as debilitate, performance.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Twenty years ago, Steele and Aronson (1995) coined the term Stereotype Threat (ST) to describe the concern that arises when one feels at risk of confirming a negative stereotype about one's group. Subsequent research has focused on how this concern debilitates the performance of stigmatized groups. For example, when examining gender stereotypes, the typical question has been, "Why do women underperform under stereotype threat?" (Cadinu, Maass, Rosabianca, & Kiesner, 2005). Consistent with the premise of this question, research has shown debilitation in a variety of cognitive and sensorimotor tasks. However, could the expectation that women underperform under ST be just another stereotype?

The focus on the debilitating effects of ST may stem from its potential negative societal implications. For example, lower ability in science and math is one of the most prominent stereotypes of females that may account for the underrepresentation of females in these fields (Eccles, Jacobs, & Harold, 1990; Nosek et al., 2009). Furthermore, Spencer, Steele, and Quinn (1999) demonstrated that women with a strong mathematical training performed worse than men with average training on the advanced GRE exam in mathematics; they performed only equally well on a comparable GRE exam of average difficulty. Critically,

when women were told that the difficult exam did not produce gender differences, they performed as well as men, suggesting that stereotypes about math ability had influenced their performance. Thus, a better understanding of ST effects may prevent the failure of young women and encourage and enable them to pursue careers in STEM fields (e.g., 28% of STEM tenure-track faculty in the US were female in 2013; National Science Foundation, National Center for Science and Engineering Statistics, 2013).

In fact, the current accounts of ST effects flow directly from this emphasis on debilitation in the cognitive domain. The prevailing perspective argues that concern over confirming the stereotype produces disrupting thoughts that utilize cognitive resources, which could be otherwise devoted to task performance. It is this reduction in working memory capacity that causes the debilitation so often reported on cognitive tasks (Schmader, Hall, & Croft, 2015; Schmader, Johns, & Forbes, 2008).

The effects of ST on motor performance have also been studied, although much less extensively than on cognitive tasks. Again, most research regarding the effect of ST on sensorimotor performance has observed debilitating effects. Studies have reported debilitation from ST in a variety of sensorimotor tasks such as golf putting (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Stone, Lynch, Sjomeling, & Darley, 1999; Stone & McWhinnie, 2008), soccer dribbling (Chalabaev, Sarrazin, Stone, & Cury, 2008; Heidrich & Chiviakowsky, 2015), simulated driving (Yeung & von Hippel, 2008), tennis serving (Hively & El-Alayli, 2014), and basketball free throw shooting (Hively & El-Alayli,

* Corresponding author at: 360 Huntington Avenue, 134 Mugar Life Sciences Building, Boston, MA 02115, USA.

E-mail address: mehuber@coe.neu.edu (M.E. Huber).

2014; Krendl, Gainsburg, & Ambady, 2012). The majority of these studies examined the effects of a gender-related ST, reporting that female performance is debilitated when exposed to the stereotype that females perform worse than males either in athletic performance or in that specific motor task (Chalabaev et al., 2008; Heidrich & Chiviawsky, 2015; Hively & El-Alayli, 2014; Stone & McWhinnie, 2008; Yeung & von Hippel, 2008). While less commonly studied, it has also been shown that male performance in golf putting can be debilitated when instructed that females perform this task better than males (Beilock et al., 2006). In addition, evoking race-related stereotypes has led to debilitated sensorimotor performance in the stigmatized group (Krendl et al., 2012; Stone et al., 1999). These reports are consistent with the pervasive stereotype that males are more competent in athletics (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002) and show higher levels of daily physical activity (Knisel, Opitz, Wossmann, & Keteihuf, 2009).

To explain the debilitating effects in motor performance, Schmader et al. (2015) suggest that ST increases performance monitoring, which in turn reduces working memory and disrupts task execution. This is particularly noticeable in well-learned, proceduralized tasks such as golf putting. However, a recent study by Huber, Seitchik, Brown, Sternad, and Harkins (2015) found that the same ST manipulation could be used to debilitate and facilitate motor performance under different circumstances. While any observation of facilitated performance under ST is incongruent with the predictions of the working memory account, the findings of Huber et al. (2015) were consistent with an alternative account developed by Jamieson and Harkins (2007). This account, referred to as “mere effort,” argues that individuals faced with ST are motivated to disprove the negative stereotype about their group, leading to the potentiation of the dominant, or prepotent, response. For sensorimotor tasks, the prepotent response is considered the dominant motor behavior, which can either be correct or incorrect, depending on whether or not the dominant motor behavior leads to the desired performance of the task. Huber et al. (2015) reported that ST affected performance in a rhythmic ball bouncing task in opposite ways, depending on the correctness of the prepotent response. This response was determined by the skill level of the performer: In novices, the prepotent response was incorrect, and therefore ST debilitated their performance; for those experienced in the task, the prepotent response was correct, and ST therefore facilitated their performance. This latter finding highlighted a largely neglected fact: under certain conditions, women may actually rise to the challenge and improve their performance under ST (e.g. Jamieson & Harkins, 2007, 2009, 2011; O'Brien & Crandall, 2003). We believe that research on facilitation under ST is very relevant, since a better understanding of how and when ST facilitates performance can also help us better understand conditions under which ST debilitates performance.

In Huber et al. (2015), facilitation due to ST was only observed for performers experienced with the task. All prior work investigating the effect of ST on motor performance for novices has reported debilitation (Heidrich & Chiviawsky, 2015; Krendl et al., 2012; Stone & McWhinnie, 2008). In the current work, we asked if ST could also facilitate the performance of inexperienced performers on a novel sensorimotor task. Following our previous results that it is the dominant behavior that determines the effect of ST, we chose a task where this dominant behavior was correct. Unlike the distinction between novices and experts, the mere effort account grounds its predictions on the correctness of the dominant or prepotent behavior. Thus, in order to observe facilitation in novices, we first had to identify a motor task where the dominant behavior was correct in novice subjects. Given the correct dominant response, the mere effort account predicted that novice performance would be facilitated. In contrast, the working memory account predicted debilitation.

2. Baseline experiment

The purpose of the baseline experiment was to identify a task where the dominant behavior of novices was correct and quantify this behavior. The experiment introduced a discrete version of the ball bouncing task, where subjects hit a ball to a target line in a single bounce. This task resembled the golf putting accuracy task frequently used in prior ST research (Beilock et al., 2006; Stone & McWhinnie, 2008; Stone et al., 1999). In aiming tasks, errors in motor performance can be caused by a constant bias (e.g., tendency to under- or overshoot the target) and/or by variability around the desired solution (Schmidt & Lee, 2005). A constant offset would suggest that the prepotent response was incorrect, whereas the absence of a bias (i.e. variability is clustered evenly around the target) would suggest that the prepotent response was correct.

It is important to note that while the experimental setup of the discrete ball bouncing task was similar to the rhythmic ball bouncing task used in our previous experiments (de Rugy, Wei, Müller, & Sternad, 2003; Dijkstra, Katsumata, de Rugy, & Sternad, 2004; Ehrlenspiel, Wei, & Sternad, 2010; Huber et al., 2015), the motor control demands were very different as different motor strategies are used in discrete versus continuous rhythmic performance (Hogan & Sternad, 2007).

2.1. Method

2.1.1. Participants

25 undergraduate students (13 males and 12 females) from Northeastern University participated in the experiment in exchange for partial fulfillment of a course requirement. None had any prior experience with the specific task. Prior to the experiment, participants read and signed the consent form as approved by the Institutional Review Board of Northeastern University. We planned to recruit an equal number of males and females, however data collection had to be terminated at the end of the semester, leading to the slightly uneven numbers.

2.1.2. Task

In the experimental task, the participants used a real racket to bounce a virtual ball to a target line (for a detailed description of the experimental setup, see Wei, Dijkstra, & Sternad, 2007). The participants stood in front of a projection screen holding a real table tennis racket in his or her dominant hand (Fig. 1). The screen displayed a virtual scene consisting of a ball, a racket, a target line positioned 1.0 m above the racket, and a number score. The vertical displacements of the real racket controlled the vertical position of the virtual racket.

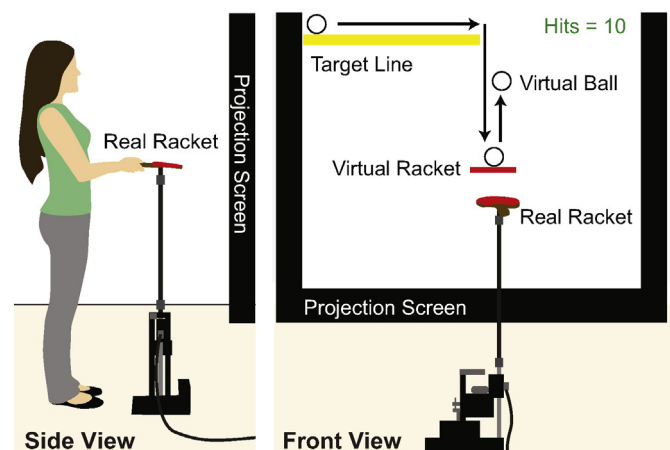


Fig. 1. Side and front view of the virtual experimental setup for discrete ball bouncing. Participants were positioned in front of a screen and manipulated a real table tennis racket to bounce a virtual ball to a target height in a 2D virtual environment.

At the start of each trial, the ball appeared at the left side of the screen and then rolled horizontally along the target line to the center of the screen (Fig. 1). Upon reaching the center, the ball dropped vertically from the target line towards the virtual racket. The participant was instructed to bounce the ball such that the maximum ball height was within ± 3 cm of the center of target line. For ball amplitudes higher or lower than this distance, the ball no longer overlapped with the target line. The vertical position of the virtual ball after ball-racket impact was determined using the equations for ballistic flight (see “Equations for Simulation of Ball Movement” in the supplementary material). Successful bounces within ± 3 cm of the target line were signaled with a temporary color change of the target line, which acted as a reward signal to the subjects. Participants were instructed to produce as many successful bounces as possible. The number of successful bounces in each block was displayed as a score on the top right corner of the screen. Following a brief pause, the next trial began. Each trial or bounce lasted approximately 3.5 s. All participants were given two practice trials for familiarization and then completed 12 blocks of 30 trials each under the watch of the experimenter, with a short break after block 6. The experiment lasted approximately 30 min.

2.1.3. Dependent measures

The first measure to characterize task performance was the percentage of successful bounces in each block of 30 trials/bounces. Bounces were deemed successful when the error was between ± 3 cm. The second measure was error, defined as the signed difference between the maximum ball height and the target line (1.0 m). The median of errors of the 30 trials/bounces in a block was calculated to serve as a measure of central tendency. Median error values outside the success region (± 3 cm) would indicate that subjects had a systematic bias or a tendency to under- or overshoot the target line. Variability, the third measure, was quantified by the interquartile range (IQR) of error in each block. IQR, the range of the second and third quartile of the distribution, is a frequently used measure to estimate dispersion. Median and IQR were used as Shapiro-Wilk tests revealed that the distributions of error were not normal in approximately 3 out of 12 blocks for each subject (Shapiro & Wilk, 1965).

2.1.4. Statistical analyses

The three dependent measures were analyzed with a 2 (Gender) \times 12 (Block) ANOVA, with gender as between-subject factors, and block as a within-subject factor. The Greenhouse-Geisser correction factor was applied to the within-subject effects (Kirk, 1995). Relevant planned comparisons using independent sample *t*-tests investigated group effects.

2.2. Results

2.2.1. Task success

The percentage of successful bounces per block was used to measure overall task performance and its improvement with practice (Fig. 2A). The ANOVA revealed a main effect for block, $F(11, 253) = 9.72$, $p < 0.001$, $\eta_p^2 = 0.30$, indicating that the percentage of successful bounces increased over the course of the 12 blocks. The ANOVA also yielded a weakly significant Gender \times Block interaction, $F(11, 253) = 2.14$, $p = 0.042$, $\eta_p^2 = 0.085$. Planned comparisons revealed that there were no significant differences between males and females on any blocks except the last block where males performed significantly better ($M = 33.33\%$, $SD = 11.94\%$) compared to females ($M = 23.89\%$, $SD = 10.72\%$), $t(23) = 2.07$, $p = 0.049$. There was no significant main effect for Gender, $F(1, 23) = 0.59$, $p > 0.250$. While participants did increase their percentage of successful bounces with practice, they never exceeded $>50\%$ success (Fig. 2A). This signaled that participants were indeed novices and that the task was challenging. Table 1 presents the overall means and standard deviations for the dependent measures for each gender.

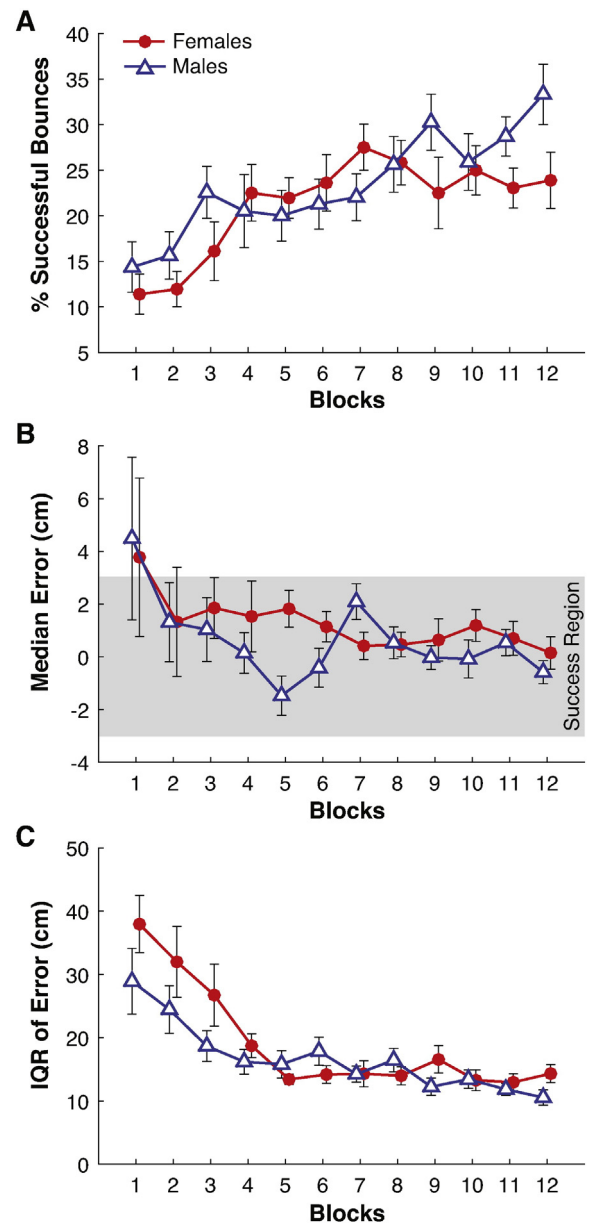


Fig. 2. Results of baseline experiment. (A) Task success, (B) median error, and (C) IQR error over blocks. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.

2.2.2. Median error

In contrast to the successful bounces, the median error showed no significant change across blocks as confirmed by the ANOVA, $F(11, 253) = 1.79$, $p = 0.162$ (Fig. 2B). All participants showed median errors that were in the gray success region, implying that they accurately centered their performance on the target line. There was no significant effect of gender, $F(1, 23) = 0.82$, $p > 0.250$, nor an interaction, $F(1,$

Table 1

Overall means and standard deviations of each dependent measure for the between-subject factor in the Baseline Experiment.

	Males M (SD)	Females M (SD)
% of Successful Bounces	23.36 (7.52)	21.27 (5.85)
Median Error (cm)	0.63 (1.80)	1.25 (1.62)
IQR of Error (cm)	16.73 (5.91)	19.03 (5.86)

23) = 0.56, $p > 0.250$ (see Table 1). This observation indicated that subjects did not have a constant bias in their motor performance. Thus, the dominant behavior, or prepotent response of novices, was correct, both in males and females.

2.2.3. Variability of error

The variability measure, IQR, quantified the distribution of errors and expressed how precisely participants hit the target line. As for the number of successful bounces, a main effect of block was revealed by the ANOVA, $F(11, 253) = 19.43$, $p < 0.001$, $\eta_p^2 = 0.46$ (Fig. 2C). All participants decreased their variability with practice. There was no significant effect of gender, $F(1, 23) = 0.96$, $p > 0.250$, nor an interaction, $F(1, 23) = 2.00$, $p = 0.132$ (Table 1). It should be noted that unlike in the measure of task success, a gender difference did not emerge in this measure, even though it is more fine-grained and has higher-resolution.

2.3. Discussion

Results showed that novices accurately centered their performance on the target line from the outset of the performance. As there was no constant offset, it was concluded that the dominant response was correct, even though the participants were all novices to the task. Performance improvements were the outcome of refining this dominant behavior by decreasing their variability and thereby increasing the percentage of successful bounces.

3. Experiment with stereotype threat

Having established the predominant behavior in the discrete bouncing task, this experiment introduced a ST manipulation. This instruction, shown to be effective in prior work, implied that females would show inferior performance in this visual-spatial task and that it was related to math ability. Importantly, extending from results in the baseline experiment, differential predictions could be made for the working memory account and the mere effort account. The mere effort account predicts that ST would potentiate the overall correct behavior and thereby improve or facilitate the performance of novices. On the other hand, the working memory account hypothesizes that novices should show inferior performance under ST, although only if the novel task “tests the upper bound of one’s skill level” (Schmader et al., 2015, p. 450). Given that subjects in the baseline experiment still performed with low success rates by the end of practice, the working memory account predicts that novices under ST show debilitated performance.

3.1. Methods

3.1.1. Participants

48 undergraduate students (24 males and 24 females) from Northeastern University participated in the experiment in exchange for partial fulfillment of a course requirement. None participated in the baseline experiment or had any prior experience with the specific task. Prior to the experiment, participants read and signed the consent form as approved by the Institutional Review Board of Northeastern University.

3.1.2. Task and ST manipulation

The task and procedure were identical to that of the baseline experiment, but with two additions. First, after completing the two practice bounces prior to the first block of trials, a male experimenter administered the following verbal instructions to participants in the ST condition:

The task you are about to complete is a test of visuo-spatial capacity. Performance on this task is closely linked to math ability. As you may know, there has been some controversy about whether there are gender differences in math and spatial ability. Previous research

has demonstrated that gender differences exist on some of these tasks, but not on others. In our lab, we examine performance on both kinds of tasks. The task on which you are about to participate has been shown to produce gender differences.

The instruction implied that males outperform females in this task, which is a negative stereotype for females. For males, this instruction presented a negative stereotype about another group. Therefore, only the females who received the manipulation experienced stereotype threat. This verbal instruction has been shown to produce ST effects in previous research (e.g. Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Spencer et al., 1999). Participants in the No-Threat (NT) condition heard the same manipulation with the sole difference that the final sentence read: “The task on which you are about to participate has not been shown to produce gender differences.”

Second, upon completion of block 12, participants filled out 11-point scales that assessed the effectiveness of the ST manipulation. The effect of the ST manipulation was gauged based on responses to two questions: (1) “To what extent do you believe that gender differences exist on this task?” (1 indicated no gender difference and 11 reflected gender differences); and (2) “Who do you believe performs better on this task?” The scale’s midpoint at 6 indicated that males and females performed the same, while lower values indicated males performed better than females and higher scores reflected the converse.

3.1.3. Statistical analyses

The same dependent measures as in the baseline experiment were analyzed in 2 (Gender) \times 2 (Threat) \times 12 (Block) ANOVAs, with gender and threat as between-subject factors, and block as a within-subject factor. The Greenhouse–Geisser correction factor was applied to the within-subject effects (Kirk, 1995). The analyses of the manipulation check excluded the block factor, but were otherwise identical.

3.2. Results

3.2.1. Manipulation check

A 2 (Gender) \times 2 (Threat) between-subjects ANOVA indicated that participants in the ST condition believed that gender differences existed to a greater extent ($M = 6.71$, $SD = 2.37$) than the participants in the NT condition ($M = 2.67$, $SD = 2.04$), $F(1, 44) = 38.46$, $p < 0.001$, $\eta_p^2 = 0.47$. Neither the gender main effect, nor the interaction was significant, $ps > 0.250$. Participants in the ST condition also reported that males outperformed females on this task to a greater extent ($M = 3.71$, $SD = 1.08$) than NT participants ($M = 5.38$, $SD = 0.92$), $F(1, 44) = 31.54$, $p < 0.001$, $\eta_p^2 = 0.42$. Again, neither the gender main effect, nor the interaction was significant, $ps > 0.250$. These results indicated that the manipulation not only successfully induced ST in females, but that the ST instruction also conveyed a positive stereotype for males.

To further confirm that there was no influence of the NT instruction on behavior, the behavior of participants in the NT condition was compared to that of the participants in the baseline condition. All dependent measures were analyzed using 2 (Gender) \times 2 (Condition) \times 12 (Block) ANOVAs, with gender and condition (NT or baseline) as between-subject factors, and block as a within-subject factor. The analyses revealed no significant main effects for condition or gender, nor any significant interactions on any measure (see “Statistical Comparisons of Baseline Participants with No-Threat (NT) Group” in the supplementary material).

3.2.2. Task success

The percentage of successful bounces per block was used to measure overall task performance (Fig. 3). The ANOVA revealed a main effect for block, $F(11, 484) = 46.27$, $p < 0.001$, $\eta_p^2 = 0.51$, indicating that the percentage of successful bounces increased over the course of the 12 blocks. Replicating the results of the baseline experiment, subjects in this

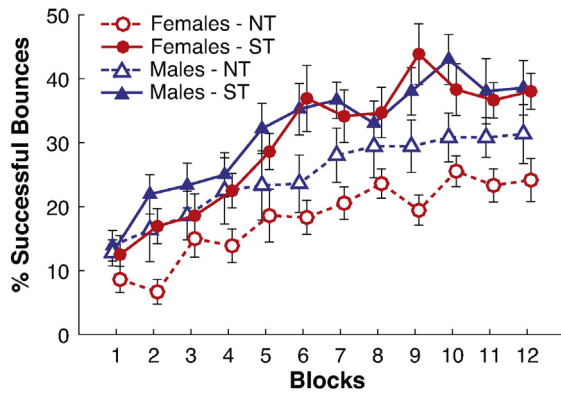


Fig. 3. Task success in ST experiment. Each point represents the group average of task success per block, and the error bar represents the standard error across participants in each group.

experiment never exceeded >50% success (Fig. 3). The analysis also revealed a main effect for threat, $F(1, 44) = 14.57$, $p < 0.001$, $\eta_p^2 = 0.25$, showing that all participants in the ST condition had a higher percentage of successful bounces ($M = 30.8\%$, $SD = 8.0\%$) than their non-threatened counterparts ($M = 21.5\%$, $SD = 9.3\%$). While both males and females in the ST condition showed better performance, it is important to note that the females acted in response to a negative stereotype to themselves, whereas males performed better upon hearing a negative stereotype about others.

The ANOVA also yielded a significant Threat \times Block interaction, $F(11, 484) = 2.43$, $p = 0.017$, $\eta_p^2 = 0.05$. While there was no significant difference between the conditions in block 1, group differences emerged over the course of the task. The main effect for gender did not reach significance, $F(1, 44) = 2.72$, $p = 0.106$, nor any of the remaining interactions, Gender \times Block, $F(11, 484) = 1.17$, $p > 0.250$, Gender \times Threat, $F(1, 44) = 0.26$, $p > 0.250$, and Gender \times Threat \times Block, $F(11, 484) = 0.58$, $p > 0.250$. Table 2 presents the overall means and SDs for each condition and gender.

3.2.3. Median and IQR of error

Fig. 4 illustrates how the distribution of error changed over blocks in two example female participants, one in the ST condition and one in the NT condition. Initially, performance was highly dispersed, but nevertheless clustered over the success region highlighted in gray for both participants. As summarized in Fig. 5A, this pattern was representative for all participants, implying that they accurately centered their performance on the target line right from the onset of practice. In accordance with this observation, the median error showed no significant change across blocks or conditions as confirmed by the ANOVA ($ps > 0.250$; Table 2). This observation indicated that the prepotent response of novices, seen already in the baseline experiment, was correct for this task and did not change with practice.

Submitting the IQR of error to the 2 (Gender) \times 2 (Threat) \times 12 (Block) ANOVA revealed a main effect for threat, $F(1, 44) = 17.70$, $p < 0.001$, $\eta_p^2 = 0.29$. Participants in the ST condition hit the ball with less variability or smaller IQR ($M = 12.6$ cm, $SD = 3.6$ cm) than the

NT participants ($M = 20.2$ cm, $SD = 8.0$ cm). This analysis also revealed a significant main effect for block, $F(11, 484) = 46.27$, $p < 0.001$, $\eta_p^2 = 0.51$. A significant Threat \times Block interaction, $F(11, 484) = 2.85$, $p = 0.021$, $\eta_p^2 = 0.06$, resulted from the fact that variability dropped significantly from block 1 to block 2, $F(1, 484) = 12.10$, $p < 0.001$, for the ST participants, but not for NT participants, $F < 1$. There was neither a main effect for gender, $F(1, 44) = 1.58$, $p = 0.216$, nor interactions of Gender \times Block, $F(11, 484) = 1.17$, $p > 0.250$, Gender \times Stereotype, $F(1, 44) = 0.26$, $p > 0.250$, and Gender \times Stereotype \times Block, $F(11, 484) = 0.21$, $p > 0.250$ (see Table 2).

To assess the robustness of the results in this experiment, we also compared the behavior of participants in the ST condition to the participants in the baseline condition (see “Statistical Analyses of Baseline Participants with Stereotype Threat (ST) Group” in the supplementary material). Overall, the results consistently demonstrated that the ST enhanced performance. While the results of this analysis revealed similar effects of ST, caution needs to be applied as we did not control for any stereotype that participants in the baseline experiment may have held.

3.3. Discussion

As in the baseline experiment, the median error was in the success region right from the outset and was not affected by the threat instruction. In contrast, the variability measure improved in all participants under the ST instruction, leading to a higher percentage of successful bounces. This facilitation, induced by the ST instruction, is consistent with the motivational mere effort account and inconsistent with the working memory account. Based on the overall correct behavior observed in the baseline experiment and corroborated in the ST experiment, the mere effort account argues that participants should be motivated by ST and perform better through potentiation of the prepotent response.

We also found that males in the ST condition outperformed their male NT counterparts. To understand this finding, it should be kept in mind that the same verbal instruction was probably perceived differently by each gender. Thus, the facilitated performance of males in the ST condition did not result from stereotype threat, but rather may have resulted from stereotype lift. We discuss this distinction further in the general discussion.

4. General discussion

Stereotype threat research has primarily focused on, and repeatedly produced, debilitated performance on cognitive and motor tasks. While this addresses a serious societal concern, we argue that under certain conditions ST may also improve performance. Therefore, any account of ST effects must give serious consideration to the possibility that ST can both facilitate and debilitate performance. With this goal, the current study examined conditions where ST could indeed facilitate performance, particularly in novices.

4.1. Eliciting threat by evoking a negative stereotype

The verbal instruction implied that males should outperform females in this discrete ball bouncing task. This same verbal instruction

Table 2

Overall means and standard deviations of each dependent measure for between-subject factors in the Stereotype Threat Experiment.

	Males		Females	
	NT <i>M (SD)</i>	ST <i>M (SD)</i>	NT <i>M (SD)</i>	ST <i>M (SD)</i>
% of Successful Bounces	24.86 (11.31)	31.51 (8.08)	18.15 (5.13)	30.16 (8.19)
Median Error (cm)	1.58 (2.45)	0.93 (2.31)	0.86 (2.94)	0.22 (2.43)
IQR of Error (cm)	18.60 (9.03)	11.97 (3.48)	21.77 (6.88)	13.31 (3.66)

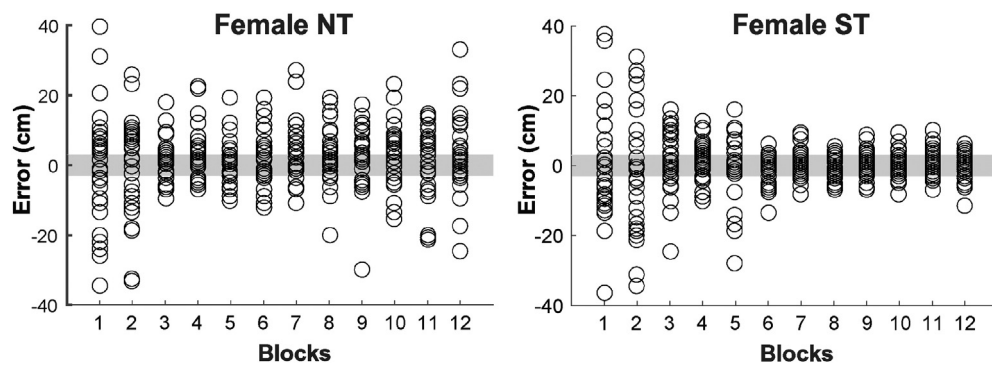


Fig. 4. Distribution of errors in each block for two example female participants from the NT group and the ST group. Each circle represents one trial or bounce. The shaded region represents errors small enough to be deemed a successful bounce (± 3 cm).

has been shown to produce ST effects in several previous studies (e.g. Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999). Huber et al. (2015) also reported that this instruction could produce both debilitation and facilitation in stigmatized females. Thus, it is safe to conclude that this instruction contains no bias that may elicit facilitation. As subjects never reached a high level of performance, it is also unlikely that the facilitation was due to a “challenge” response to the stereotype instruction. Such a response occurs when subjects believe their abilities are sufficient to meet the task demands. It differs from a “threat” response, where subjects assess their abilities as insufficient for accomplishing the task (Blascovich & Mendes, 2000). Based on these considerations, we posit that the facilitated performance of stigmatized females was in fact due to stereotype threat.

4.2. Effect of stereotype threat on motor variability

Before discussing the theoretical implications of the findings from the ST experiment, it is important to highlight that this facilitating effect was evidenced by a reduced variability in the ST group, not by a change in the central tendency (i.e. mean or median) as typically considered. In basic motor control research, variability in task performance has been widely recognized as an essential window into control processes (Davids, Bennett, & Newell, 2006; Newell & Corcos, 1993; Sternad, Abe, Hu, & Müller, 2011). Variability is a measure independent from average behavior and analyses of its temporal and distributional structure have revealed numerous insights into sensorimotor control (Abe & Sternad, 2013; Cohen & Sternad, 2009; Gilden, Thornton, & Mallon, 1995; Sternad, Dean, & Newell, 2000). Unlike in cognitive tasks, where the verbal or written output directly reflects the response from executive centers, motor performance is subject to many additional peripheral and central processes that add variability to the initial plan. Hence, measures of variability present informative variables that have not yet been exploited in studies of ST on performance.

4.3. Theoretical implications of experimental results

The goal of the study was to compare the predictions of the working memory and mere effort accounts on sensorimotor performance of female novices under ST. The working memory account predicted debilitation for novices in a challenging task. In contrast, the mere effort account predicted facilitation, if the predominant behavior was regarded correct. Creating a task scenario that was challenging but nevertheless showed a correct behavior even in novices, participants under ST were shown to improve their performance by decreasing their variability and increasing their successful target hits.

Schmader et al. (2015) suggest that ST increases performance monitoring, which in turn reduces working memory and disrupts task execution on well-learned, proceduralized tasks. However another account, the explicit monitoring account, proposes different underlying processes. It suggests that it is the availability of working memory, rather than its absence, that produces this debilitation on well-learned tasks (Beilock et al., 2006). The authors propose that ST causes experts to attend closely to sensorimotor processes and this explicit attention to step-by-step processing is thought to disrupt task execution that normally runs outside of conscious awareness (Baumeister, 1984; Beilock, Bertenthal, McCoy, & Carr, 2004; Gray, 2004; Langer & Imber, 1979; Masters, 1992). If ST prompts explicit monitoring as suggested, it then follows that for novices, ST-induced explicit monitoring should facilitate the performance (Beilock, Carr, MacMahon, & Starkes, 2002; Beilock et al., 2004; DeCaro, Thomas, Albert, & Beilock, 2011; Gray, 2004). The results of this experiment support these predictions. However, the explicit monitoring account does not explain the debilitation that is, in fact,

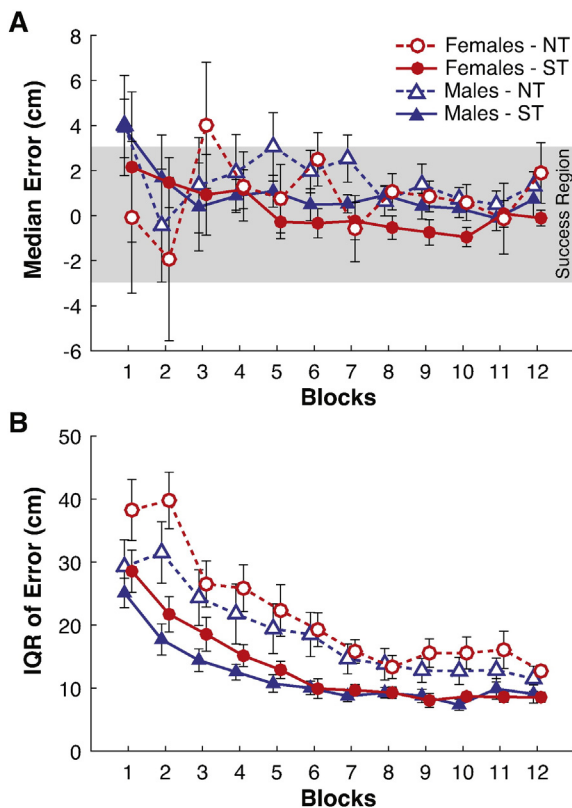


Fig. 5. (A) Median and (B) IQR of error in ST experiment. Each point represents the group average of the dependent measures per block, and the error bar represents the standard error across participants in each group.

more commonly observed in stigmatized novices. Recently, Chalabaev and colleagues also questioned both working memory and explicit monitoring accounts, as they observed that ST debilitated performance on a ballistic isometric force task, even though the task was too short to allow any explicit monitoring processes relying on working memory (Chalabaev et al., 2013).

The mere effort account explains the present findings and also other experimental results, previously reported as support for the explicit monitoring account. For example, Beilock et al. (2006) showed that ST debilitated expert players in a golf putting task and attributed this performance decrement to undue attention to processes that usually run automatically. While plausible, their findings can also be explained by the mere effort account. In their task, golf experts were asked to putt the ball so that it stopped directly on the target. This requirement differed from how these experts typically putt, making the task more difficult. Putting the ball through the target and overshooting is the most typical strategy for expert golfers, which was the incorrect behavior for this experiment. Hence, the reported debilitation would be consistent with the mere effort account.

4.4. Novice and expert performance versus correct and incorrect behavior

Unlike the other accounts, mere effort account does not distinguish between novices and experts, rather, between correct and incorrect prepotent behavior. In the rhythmic ball bouncing task and in a discrete golf putting task, ST debilitated novice performance (Huber et al., 2015; Stone & McWhinnie, 2008), whereas ST facilitated the performance of stigmatized novices in this discrete ball bouncing task. These seemingly contradictory results raise the possibility that the distinction between “novice” and “expert” may not be the most appropriate to explain ST effect on sensorimotor performance (Chalabaev et al., 2008; Stone & McWhinnie, 2008). The notion that performance under threat is contingent upon the correctness of the prepotent response is consistent with other accounts that suggested that the effect of ST depends on whether the task was “easy” or “hard” (O’Brien & Crandall, 2003) or whether the task is “well-learned” or not (Beilock et al., 2006). While tasks characterized as “easy” often have correct prepotent responses, this is not always the case. Harkins (2006) showed that prepotency precedes task difficulty when predicting performance outcomes for stigmatized individuals. The same is true of the relationship between task experience and prepotency (i.e., experts do not always have a correct prepotent response). Furthermore, there is no consensus or quantitative criteria for such categorizations. Evaluating the behavior rather than the task offers a more principled criterion, as well as more effective means to making predictions for performance under threat.

4.5. Stereotype lift in males

A secondary finding was that the male subjects who received the same verbal instruction also outperformed their NT counterparts. This behavior is consistent with the notion of stereotype lift, previously observed in males in ST research (e.g. Chalabaev et al., 2013, 2008; Chatard, Selimbegović, Konan, & Mugny, 2008; Croizet et al., 2004; Laurin, 2013; Shih, Ambady, Richeson, Fujita, & Gray, 2002; Shih, Pittinsky, & Ambady, 1999; Walton & Cohen, 2003; Wraga, Helt, Jacobs, & Sullivan, 2007). Stereotype lift has been suggested to increase motivation to uphold the manipulated stereotype and buttress self-esteem. While the behavioral results show similar decreases in variability for both genders in the ST condition, these results are likely to have occurred through different processes. The same verbal instruction conveyed very different messages to the participants: for example, males may have been motivated to confirm the negative stereotype about females, and by extension, confirm a positive stereotype about males. In contrast, females are motivated to disconfirm the negative stereotype. Indeed, other research on sensorimotor performance, activating

negative stereotypes about females led males to perform better through increases in effort (Chalabaev, Stone, Sarrazin, & Croizet, 2008).

4.6. Study limitations

While the experimental results provide support for the mere effort account, the study also had some limitations. A first caveat is the identification of the prepotent response. While we could identify the prepotent response through behavioral data, it is not possible to extract the prepotent response at the neural level. Second, this task could not test the mere effort account's prediction of debilitation. However, in the previous study that used the rhythmic version of this task (Huber et al., 2015), both facilitation and debilitation was predicted and observed. Note, though, that performance measures of rhythmic and discrete execution are different. Hence, an interesting experimental test would be to use the discrete task variant and teach subjects the wrong prepotent response to then test both predictions.

One additional limitation was that we only examined the effect of ST on the initial performance of novices. It is just as important to know whether the effect of ST would persist if novices continued to improve task performance over several days of practice. Nevertheless, the results of this study clearly mark the need for a more comprehensive explanation of ST, including facilitation under several different circumstances.

5. Conclusions and outlook

While a great deal of research has focused on explaining how ST debilitates performance, any comprehensive account should also be able to explain how ST can facilitate performance. Not only do we observe facilitation under ST in the lab, but we also see women surpass men at elite sports despite the still prevailing female stereotype in athletics. Famous examples are boxer Jackie Tonawanda and the bowler Kelly Kulick. Perhaps by understanding the conditions in which ST improves performance, we may gain insights on how to help females perform better in the presence of pervasive societal stereotypes. The results of this study also underscored that measures of motor variability are essential to understand how subtle psychological manipulations affect performance in a sensorimotor task. Hence, we agree with the editors of a recent handbook on ST who referred to their volume as only the “halftime report” (Inzlicht & Schmader, 2012) and argue that a more comprehensive account of ST is required.

Disclosure statement

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

Acknowledgement

This work was supported by U.S. Army Research Institute for the Behavioral and Social Sciences under Contract W5J9CQ-12-C-0046; the National Institutes of Health under Grant R01-HD045639; and the National Science Foundation under Grant DMS-0928587. The views, opinions, and/or findings are those of the authors and not an official Department of the Army position, policy, or decision. We would like to thank S.G. Harkins for his assistance in this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.actpsy.2016.05.010>.

References

- Abe, M. O., & Sternad, D. (2013). Directionality in distribution and temporal structure of variability in skill acquisition. *Frontiers in Human Neuroscience*, 7, 225. <http://dx.doi.org/10.3389/fnhum.2013.00225>.
- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610–620. <http://dx.doi.org/10.1037/0022-3514.46.3.610>.
- Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, 8(1), 6–16. <http://dx.doi.org/10.1037/1076-898X.8.1.6>.
- Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review*, 11(2), 373–379. <http://dx.doi.org/10.3758/BF03196585>.
- Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin*, 32(8), 1059–1071. <http://dx.doi.org/10.1177/0146167206288489>.
- Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 59–82). Cambridge UK: Cambridge University Press.
- Brown, R. P., & Piel, 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](http://dx.doi.org/10.1016/S0022-1031(03)00039-8).
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, 16(7), 572–578. <http://dx.doi.org/10.1111/j.0956-7976.2005.01577>.
- Chalabaev, A., Sarrazin, P., Stone, J., & Cury, F. (2008). Do achievement goals mediate stereotype threat? An investigation on females' soccer performance. *Journal of Sport & Exercise Psychology*, 30(2), 143–158.
- Chalabaev, A., Brisswalter, J., Radel, R., Coombes, S. A., Easthope, C., & Clement-Guillotin, C. (2013). Can stereotype threat affect motor performance in the absence of explicit monitoring processes? Evidence using a strength task. *Journal of Sport & Exercise Psychology*, 35, 211–215.
- Chalabaev, A., Stone, J., Sarrazin, P., & Croizet, J. (2008). Investigating physiological and self-reported mediators of stereotype lift effects on a motor task. *Basic and Applied Social Psychology*, 30(1), 18–26. <http://dx.doi.org/10.1080/01973530701665256>.
- Chatard, A., Selimbegović, L., Konan, P., & Mugny, G. (2008). Performance boosts in the classroom: Stereotype endorsement and prejudice moderate stereotype lift. *Journal of Experimental Social Psychology*, 44(5), 1421–1424. <http://dx.doi.org/10.1016/j.jesp.2008.05.004>.
- Cohen, R. G., & Sternad, D. (2009). Variability in motor learning: Relocating, channeling and reducing noise. *Experimental Brain Research*, 193(1), 69–83. <http://dx.doi.org/10.1007/s00221-008-1596-1>.
- Croizet, J. C., Després, G., Gauzins, M. E., Huguet, P., Leyens, J. P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, 30(6), 721–731. <http://dx.doi.org/10.1177/0146167204263961>.
- Davids, K., Bennett, S., & Newell, K. (Eds.). (2006). *Movement system variability*. Champaign, IL: Human Kinetics.
- de Rugy, A., Wei, K., Müller, H., & Sternad, D. (2003). Actively tracking 'passive' stability in a ball bouncing task. *Brain Research*, 982, 64–78. [http://dx.doi.org/10.1016/S0006-8993\(03\)02976-7](http://dx.doi.org/10.1016/S0006-8993(03)02976-7).
- DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General*, 140(3), 390–406. <http://dx.doi.org/10.1037/a0023466>.
- Dijkstra, T. M. H., Katsumata, H., de Rugy, A., & Sternad, D. (2004). The dialogue between data and model: Passive stability and relaxation behavior in a ball bouncing task. *Nonlinear Studies*, 11, 319–344.
- Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, 46(2), 183–201.
- Ehrlenspiel, F., Wei, K., & Sternad, D. (2010). Open-loop, closed-loop and compensatory control: Performance improvement under pressure in a rhythmic task. *Experimental Brain Research*, 201(4), 729–741. <http://dx.doi.org/10.1007/s00221-009-2087-8>.
- Gilden, H. L., Thornton, T., & Mallon, M. W. (1995). 1/f in cognition. *Science*, 267(5205), 1837–1839.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, 10(1), 42–54. <http://dx.doi.org/10.1037/1076-898X.10.1.42>.
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, 91(3), 436–455. <http://dx.doi.org/10.1037/0022-3514.91.3.436>.
- Heidrich, C., & Chiviacowsky, S. (2015). Stereotype threat affects the learning of sport motor skills. *Psychology of Sport and Exercise*, 18, 42–46. <http://dx.doi.org/10.1016/j.psychsport.2014.12.002>.
- Hively, K., & El-Alayli, A. (2014). "You throw like a girl." The effect of stereotype threat on women's athletic performance and gender stereotypes. *Psychology of Sport and Exercise*, 15, 48–55. <http://dx.doi.org/10.1016/j.psychsport.2013.09.001>.
- Hogan, N., & Sternad, D. (2007). On rhythmic and discrete movements: Reflections, definitions and implications for motor control. *Experimental Brain Research*, 181(1), 13–30. <http://dx.doi.org/10.1007/s00221-007-0899-y>.
- Huber, M. E., Seitchik, A., 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(2), 525–541. <http://dx.doi.org/10.1037/xhp0000039>.
- Inzlicht, M., & Schmader, T. (Eds.). (2012). *Stereotype threat: Theory, process, and application*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199732449.001.0001>.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73(2), 509–527. <http://dx.doi.org/10.1111/1467-8624.00421>.
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93(4), 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(10), 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(5), 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(3), 371–381. <http://dx.doi.org/10.1177/0146167202250218>.
- Kirk, R. (1995). *Experimental design*. Pacific Grove, CA: Brooks/Cole.
- Knisel, E., Opitz, S., Wossmann, M., & Keteihuf, K. (2009). Sport motivation and physical activity of students in three European schools. *International Journal of Physical Education*, 46(2).
- Krendl, A., Gainsburg, I., & Ambady, N. (2012). The effects of stereotypes and observer pressure on athletic performance. *Journal of Sport & Exercise Psychology*, 34(1), 3–15.
- Langer, E. J., & Imber, L. G. (1979). When practice makes imperfect: Debilitating effects of overlearning. *Journal of Personality and Social Psychology*, 37(11), 2014–2024. <http://dx.doi.org/10.1037/0022-3514.37.11.2014>.
- Laurin, R. (2013). Stereotype threat and lift effects in motor task performance: The mediating role of somatic and cognitive anxiety. *The Journal of Social Psychology*, 153(6), 687–699. <http://dx.doi.org/10.1080/00224545.2013.821098>.
- Masters, R. S. W. (1992). Knowledge, nerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology*, 83(3), 343–358. <http://dx.doi.org/10.1111/j.2044-8295.1992.tb02446.x>.
- National Science Foundation, National Center for Science and Engineering Statistics (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013*. Arlington, VA: Special Report NSF 13-304 Available at <http://www.nsf.gov/statistics/wmpd/>.
- Newell, K. M., & Corcos, D. M. (1993). *Variability and motor control*. Champaign, IL: Human Kinetics.
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10593–10597. <http://dx.doi.org/10.1073/pnas.0809921106>.
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29(6), 782–789. <http://dx.doi.org/10.1177/0146167203029006010>.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–356. <http://dx.doi.org/10.1037/0033-295X.115.2.336>.
- 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, D.C.: American Psychological Association.
- Schmidt, R. A., & Lee, T. D. (2005). Motor control and learning: A behavioral emphasis. (5th ed.). Champaign, IL: Human Kinetics.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <http://dx.doi.org/10.2307/2333709>.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10, 80–83. <http://dx.doi.org/10.1111/1467-9280.00111>.
- Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83(3), 638–647. <http://dx.doi.org/10.1037/0022-3514.83.3.638>.
- 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(5), 797–811. <http://dx.doi.org/10.1037/0022-3514.69.5.797>.
- Sternad, D., Dean, W. J., & Newell, K. M. (2000). Force and timing variability in rhythmic unimanual tapping. *Journal of Motor Behavior*, 32(3), 249–267. <http://dx.doi.org/10.1080/00222890009601376>.
- Sternad, D., Abe, M. O., Hu, X., & Müller, H. (2011). Neuromotor noise, error tolerance and velocity-dependent costs in skilled performance. *PLoS Computational Biology*, 7(9), e1002159. <http://dx.doi.org/10.1371/journal.pcbi.1002159>.
- Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, 44(2), 445–452. <http://dx.doi.org/10.1016/j.jesp.2007.02.006>.

- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77(6), 1213–1227. <http://dx.doi.org/10.1037/0022-3514.77.6.1213>.
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39, 456–467. [http://dx.doi.org/10.1016/S0022-1031\(03\)00019-2](http://dx.doi.org/10.1016/S0022-1031(03)00019-2).
- Wei, K., Dijkstra, T. M. H., & Sternad, D. (2007). Passive stability and active control in a rhythmic task. *Journal of Neurophysiology*, 98(5), 2633–2646. <http://dx.doi.org/10.1152/jn.00742.2007>.
- Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience*, 2(1), 12–19. <http://dx.doi.org/10.1093/scan/nsl041>.
- Yeung, N. C. J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis & Prevention*, 40, 667–674. <http://dx.doi.org/10.1016/j.aap.2007.09.003>.