

Priming performance-related concerns induces task-related mind-wandering

By: Megan L. Jordano, [Dayna R. Touron](#)*

Jordano, M. L. & Touron, D. R. (2017). Priming performance-related concerns induces task-related mind-wandering. *Consciousness & Cognition*, 55, 126-135.

***© Elsevier, Inc. Reprinted with permission. No further reproduction is authorized without written permission from Elsevier, Inc. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. ***



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

Made available courtesy of Elsevier: <https://doi.org/10.1016/j.concog.2017.08.002>

Abstract:

Two experiments tested the hypothesis that priming of performance-related concerns would (1) increase the frequency of task-related mind-wandering (i.e., task-related interference; TRI) and (2) decrease task performance. In each experiment, sixty female participants completed an operation span task (OSPAN) containing thought content probes. The task was framed as a math task for those in a condition primed for math-related stereotype threat and as a memory task for those in a control condition. In both studies, women whose performance-related concerns were primed via stereotype threat reported more TRI than women in the control. The second experiment used a more challenging OSPAN task and stereotype primed women also had lower math accuracy than controls. These results support the “control failures × current concerns” framework of mind-wandering, which posits that the degree to which the environmental context triggers personal concerns influences both mind-wandering frequency and content.

Keywords: Mind wandering | Stereotype threat | Executive control

Article:

* The University of North Carolina at Greensboro, Department of Psychology, 296 Eberhart Building, PO Box 26170, Greensboro, NC 27412, United States

1. Introduction

Everyone has experienced starting a task, such as reading a journal article, only to find later that their thoughts have drifted to something else. Mind-wandering can involve intrusive thoughts that are somewhat related to the current task (e.g., “This article is confusing. I’m struggling to understand it.”), which is known as task-related interference (TRI). Other times the intrusive thoughts are completely unrelated to the task (e.g., “I’m getting hungry. What should I make for dinner?”), which are called task-unrelated thoughts (TUTs). TUTs have been the primary focus of mind-wandering research whereas mind-wandering about task performance is less well-studied (McVay, Meier, Touron, & Kane, 2013). It has been proposed that the degree to which one’s environmental context triggers one’s current concerns influences the *frequency* and *content* of one’s mind-wandering experiences (McVay & Kane, 2010). The current studies examine the effect of priming personal, performance-related concerns on mind-wandering (particularly TRI) as well as task performance.

2. Variations in frequency and content of mind-wandering

Mind-wandering occurs frequently with younger adults, on average, spending one third to one half of their daily lives thinking about something other than their current task (Kane et al., 2007; Klinger & Cox, 1987-1988; McVay, Kane, & Kwapil, 2009). Mind-wandering is even more frequent during laboratory tasks (Jackson & Balota, 2012; McVay et al., 2013). The frequency of mind-wandering varies among individuals, with individuals with lower executive control abilities mind-wandering more often (Kane et al., 2007; McVay & Kane, 2009, 2012a, 2012b) than those with higher executive control abilities.

There are contrasting theories explaining the relationship between mind-wandering and executive control. One perspective posits that mind-wandering requires executive resources (Smallwood & Schooler, 2006). Individuals experience fewer TUTs during tasks with greater cognitive loads (Antrobus, 1968; Teasdale et al., 1995). Additionally, in-the-moment TUT reports predict performance errors on the ongoing task (Smallwood et al., 2004) and individual differences in executive control abilities are positively correlated with TUT frequency on simple tasks (Levinson, Smallwood, & Davidson, 2012). This suggests that individuals with more executive control resources use their additional resources to engage in mind-wandering.

However, other theories of mind-wandering propose that we use our executive resources to prevent mind-wandering experiences. According to the “control failures x current concerns” framework of mind-wandering proposed by McVay and Kane (2010), off-task thoughts are continuously and automatically generated in response to cues in one’s environment, and we use executive control resources to prevent these off-task thoughts from entering into conscious awareness. Mind-wandering is therefore believed to result from an interplay between one’s executive control abilities, one’s motivation to prevent off-task thoughts from entering consciousness, and the degree to which the environmental context primes one’s current concerns. There is evidence supporting this perspective. Individual differences in TUT frequency are positively correlated with control abilities on simple tasks (Levinson et al., 2012). Furthermore, frequency of TUTs predicts performance deficits on more attention-demanding tasks (McVay & Kane, 2009; McVay et al., 2009; Smallwood et al., 2004) as well as executive control attempts

on ongoing, everyday tasks (Kane et al., 2017). Finally, individuals with greater executive control suffer as much as those with lower control in terms of task performance when they engage in mind-wandering (McVay & Kane, 2012a, 2012b). These outcomes suggest that, rather than using our executive control resources to engage in mind-wandering, we may instead use them to try to prevent mind-wandering when the ongoing task is difficult.

In addition to offering predictions regarding the frequency of mind-wandering, the “control failures x current concerns” framework can also predict the content of mind-wandering experiences. As mentioned, mind-wandering can be about things completely unrelated to the task (TUTs) or about concerns that are somewhat related to the task, such as evaluating one’s task performance (TRI). TRI differs conceptually from TUTs because they are task-related. However, TRI experiences are not quite on-task or directly about responding appropriately to the task stimuli and task demands. Although TRI is distinguishable from TUTs, many researchers consider TRI to be a variety of mind-wandering, given that TRI and TUTs are similarly associated with in-the-moments performance deficits (McVay & Kane, 2009; McVay et al., 2013; Mrazek et al., 2011; Smallwood, O’Connor, & Heim, 2005). Research conducted on the default mode network (DMN), a brain network that is active when individuals are at rest, has revealed that TUTs are associated with higher levels of DMN recruitment and on-task thoughts with lower levels of DMN recruitment. While TRI is not associated with as much DMN recruitment as TUTs, both TRI and distraction due to external environmental stimuli are associated with more DMN recruitment than on-task thoughts (Gonçalves et al., 2017). Given these findings, we also characterize TRI as a type of mind-wandering. TRI has been studied less frequently than TUTs and the thought probes used in many mind-wandering studies do not include TRI as a response option. Participants in these studies may therefore misclassify their TRI experiences as being on-task. In laboratory studies that have included TRI as a thought probe response option, younger adults report more TUTs, relative to both their TRI and TUTs by older adults, whereas older adults report more TRI, relative to both their TUTs and TRI by younger adults (Frank, Nara, Zavagnin, Touron, & Kane, 2015; Giambra, 1989; Grodsky & Giambra, 1990; Jackson & Balota, 2012; Krawietz, Tamplin, & Radvansky, 2012; McVay & Kane, 2012a; McVay et al., 2013). While the typical academic laboratory testing environment may cue everyday school-related concerns in younger adults (resulting in increased TUTs) this same testing environment may instead trigger concerns about cognitive decline in older adults (resulting in increased TRI). Therefore, the finding that younger adults report more TUTs and older adults report more TRI corroborates the perspective that one’s current concerns and the extent to which they are primed by the environmental context determines frequency and content of one’s mind-wandering experiences.

In the current studies, we provide a direct test of the “control failures x current concerns” framework by priming *performance-related concerns* in younger adults and assessing the impact on probe-caught off-task thoughts regarding the task (TRI). More specifically, we prime current, performance-related concerns in these younger adults using a stereotype threat intervention.

3. Stereotype threat and intrusive thoughts

Stereotype threat, or risk for behaving in a way that confirms a negative stereotype about a group one belongs to Steele & Aronson (1995), has been linked to performance deficits on a variety of

tasks in various marginalized groups. Different mechanisms have been proposed to explain the performance impairments typically observed in those under threat (for a review, see Schmader, Johns, & Forbes (2008)). These mechanisms include increased physiological stress, increased regulation of negative emotions, and increased monitoring of task performance. If a person under threat uses their resources to engage in task performance monitoring and active suppression of negative emotions (Schmader, Forbes, Zhang, & Mendes, 2009; Schmader et al., 2008; Schuster, Martiny, & Schmader, 2015), then fewer resources are available to respond appropriately to task stimuli and demands, resulting in impaired performance (Beilock, Rydell, & McConnell, 2007; Cadinu, Maass, Rosabianca, & Kiesner, 2005).

Although it has been proposed that stereotype threat increases task-related worries, research on the topic has been limited by the methodologies used to measure off-task thoughts. For example, previous work on stereotype threat and off-task thoughts has relied primarily on open verbal reports of intrusive thoughts. Coding schemes for these verbal reports may not include both TUTs and TRI (Cadinu et al., 2005) or may rely on participants thinking back and reporting on their mind-wandering episodes after the testing session is completed (Beilock et al., 2007). Retrospective mind-wandering assessments may result in mind-wandering episodes being systematically overlooked, because individuals may not always recognize or remember when their thoughts have drifted (Schooler, 2002; Smallwood, McSpadden, & Schooler, 2008).

Mrazek et al. (2011) examined the effect of math-gender stereotype threat on mind-wandering using *online thought probes* during a challenging math task. However, the methodology of this study has limited application to the current questions. The online thought probes used to assess mind-wandering asked participants to indicate, on a Likert-scale from 1 to 10, the degree to which participants were mind-wandering immediately before the probe appeared, and did not include different categories of mind-wandering for participants to choose from. Retrospective thought content assessments were administered after the testing session. A simultaneous regression predicting probe-caught mind-wandering from retrospective TUTs and retrospective TRI revealed that the online thought-probes were more strongly related to retrospective TUTs rather than retrospective TRI, although past research suggests that threat leads to increased worry-laden task evaluation of performance, which corresponds to TRI (Beilock et al., 2007; Cadinu et al., 2005). These results are dependent on participants' retrospective memories of what they were mind-wandering about during the testing session.

The current studies examine the effect of priming personal, performance-related concerns on mind-wandering by using a stereotype threat manipulation. Mind-wandering is assessed during a task with a mathematical component following priming of math-gender stereotype. Critically, mind wandering is measured using online-thought probes embedded in the task that provide participants with a variety of relevant thought content options, including TRI. Both task performance (number of errors and response time) and eye movement data corroborate online thought probes as a measure of mind-wandering (Frank et al., 2015). This thought probe approach will not only allow a test of the “control failures × current concerns” framework of mind-wandering (McVay & Kane, 2010), but will also provide a better understanding of what types of intrusive thoughts individuals under stereotype threat experience. We predict that stereotype threat priming will lead to increased worry-laden monitoring of performance, reflected in increased TRI. Furthermore, we predict that individuals under stereotype threat will

have worse performance on a task in the stereotyped domain than individuals who are not systematically primed for stereotype threat.

4. Experimental overview

To test these hypotheses, two experiments were completed. Online thought probes were used to measure mind-wandering frequency and content. Study 1 tested the effect of priming math-gender stereotype threat on mind-wandering and math verification accuracy during an automated Operation Span Task (OSPAN; Conway et al., 2005). Study 2 provides a theoretical replication of Study 1 using a more challenging version of the OSPAN. These studies, using more precise online thought probes, provide us with the opportunity to examine how current concern priming affects mind-wandering and task performance, and to discern whether TRI is partially responsible for the harmful effects of stereotype threat on task performance.

5. Study 1 method

5.1. Participants

Sixty female undergraduate students from the University of North Carolina at Greensboro participated in exchange for course credit (M age = 19.10, SD = 1.24). Half of the participants were tested in the stereotype threat condition, and the other half in the control condition. All of these participants were included in the analyses described below. We did not want participants to suspect that the current study was about stereotype threat or gender differences in cognition. Therefore we also recruited and tested male participants. Male participants were tested in both the control and stereotype threat conditions together with the female participants, but only the data of female participants were included in the following analyses.¹

5.2. Design and procedure

All measures and experimental manipulations for this study are reported below.² Study 1 was a between-subjects design comparing the frequency of mind-wandering among female undergraduates primed for stereotype threat relative to women in a control condition. Because the goal of the current study was to examine the effects of priming performance-related concerns on mind-wandering, we wanted to make sure we successfully primed stereotype threat in the experimental group. To do this, we primed stereotype threat in a couple different ways.

Participants in the experimental condition were told that the OSPAN task they were about to complete was a task that measures “quantitative ability” and that the task had revealed gender differences in the past. These participants were also tested by a male experimenter. The majority of control and stereotype threat sessions contained at least one male participant. We had a male confederate act as a participant in the stereotype threat condition sessions in the event that no male students signed up for that session. In the control condition, participants were told that the OSPAN task they were about to complete measured “memory ability” and that this task was gender-fair. Control participants were tested by a female experimenter. Participants in both groups were told that they would receive feedback for their performance.

Before completing the OSPAN participants completed a modified Gender-Science Implicit Associations Task (IAT; Nosek & Smyth, 2007). This IAT was designed to measure the participants' automatic associations between the concepts of "math" and "liberal arts" and gender. This measure was included as a manipulation check. If we had not found the anticipated group differences in mind-wandering or task performance, we could use these IAT scores to further evaluate the effectiveness of our stereotype threat priming procedure.³

After instruction, participants completed a modified, automatic OSPAN. The OSPAN is a measure of working memory capacity that requires participants to switch back and forth between verifying relatively simple math equations and viewing letters that must be remembered for a later recall task. At the end of each OSPAN trial participants must, in correct serial order, recall the letters presented to them during the trial. The OSPAN for Study 1 consisted of 130 trials with set sizes (numbers of consecutive equation/letter pairs) varying between 2 and 3. Although the OSPAN is a measure of working memory, we chose it as our main experimental task because it could be framed as either a math task or as a gender-fair memory task, depending on the experimental condition, providing us with stereotype threat manipulation. Both math verification and letter recall accuracy were recorded during the testing session.

During the OSPAN participants responded to a total of 10 thought probes which appeared at quasi-random intervals (spacing was approximately every 2 min; see Seli, Carriere, Levene, & Smilek, 2013). Probes could appear (1) after participants saw the final math equation within a trial, but before they verified the answer for that equation, (2) after the final math equation within the trial had been verified, but before the final letter to be remembered was presented, or (3) at the end of the trial after the participant saw the last letter to be remembered. OSPAN trials in which a thought probe appeared were dropped from the OSPAN analyses reported below. Participants responded to the probes using their computer keypad and were instructed to indicate what they were thinking about *immediately before* the probe appeared:

What were you just thinking about?

- (1) The task: Focused on completing the task, verifying equations, and remembering letters
- (2) Task approach: Thinking about how you can improve your task performance
- (3) Task evaluation: Evaluating how effective you were completing the task, or worrying about task performance
- (4) Everyday things: Thinking about recent or impending life events
- (5) Current state of being: Thinking about conditions such as hunger or sleepiness
- (6) Personal worries: Thinking about concerns, troubles, or fears not related to the experimental task
- (7) Daydreams: Fantasies disconnected from reality
- (8) Other.

Using these online thought probes we calculated the mean proportions of probe caught TUT, TRI, and on-task thoughts. In the current study, we also separated TRI into two different subtypes. It has been suggested that ST primarily increases worry-laden monitoring of task performance (Schmader et al., 2008), which we label "reactive TRI", corresponding to Option 3 in the thought probe example above. However, we believe it is also possible that TRI can also be

about task approach or strategy, which we label “proactive TRI”, corresponding to Option 2 above. Furthermore, we believed this proactive form of mind-wandering about task approach might have downstream benefits in terms of task performance, even if experiencing these thoughts harm performance in the moment.

In addition to online thought probes, we included a retrospective measure of thought content. The Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gililand, Huggins, & Falconer, 1999), includes eight questions that assess TUTs (e.g. “I thought about something that happened to me earlier”) and eight questions that assess TRI (e.g. “I thought about how I should work more carefully”). The DSSQ has been used in other mind-wandering studies (Mrazek et al. (2011)) and was included here to see if it corroborated the results of the online thought probes.

In addition to the TUT and TRI subscales, the Motivation subscale of the DSSQ was also included. The Motivation subscale consists of seven questions that measure motivation to do well on the experimental task (the OSPAN), along with seven items that measure participants’ interest in the content of the experimental task. The Positive and Negative Affect Scale (The PANAS; Watson, Clark, & Tellegan, 1988) was included as a measure of mood, as mood may be expected to influence the frequency of mind-wandering (Kane et al., 2007; Smallwood, Fitzgerald, Miles, & Phillips, 2009). We also included several post-task questions about perceived task difficulty, perceived fatigue during the OSPAN, and perceived stress during the OSPAN, as these may also be expected to influence frequency of mind wandering (Kane et al., 2007). To assess possible reactivity, we asked participants what they thought the study was about after they completed all tasks. No participants indicated believing that the study was about stereotype threat or gender differences in cognition.

6. Study 1 results and discussion

6.1. Performance and mind-wandering

Table 1 contains means and standard errors for OSPAN and post-task measures. As shown in Figure. 1, the proportion of probe-caught TRI was higher in stereotype threat primed participants ($M = 0.210$, $SE = 0.026$) than in control participants ($M = 0.127$, $SE = 0.023$), $F(1, 58) = 5.669$, $p = 0.021$, $d = 0.636$. Participants primed for stereotype threat reported more intrusive off-task thoughts about task performance and task strategy than participants who were not primed for threat. Likewise, participants in the control condition were more likely to indicate being on-task ($M = 0.463$, $SE = 0.041$) than participants primed for stereotype threat ($M = 0.370$, $SE = 0.041$), although this difference did not reach statistical significance ($F(1, 58) = 2.694$, $p = 0.106$).

Table 1
Study 1 and study 2 descriptive statistics.

Variable	Study 1					Study 2				
	Control		Stereotype Threat			Control		Stereotype Threat		
	M	SE	M	SE	p-value	M	SE	M	SE	p-value
OT	0.463	0.041	0.370	0.041	0.106	0.541	0.043	0.352	0.051	0.007
TUT	0.390	0.041	0.413	0.041	0.691	0.230	0.024	0.304	0.049	0.182
TRI	0.127	0.023	0.210	0.026	0.021	0.080	0.032	0.161	0.035	0.022
Proactive TRI	0.040	0.010	0.097	0.023	0.027	0.034	0.009	0.052	0.009	0.750
Reactive TRI	0.087	0.017	0.123	0.020	0.171	0.027	0.006	0.057	0.008	0.064
Math Accuracy	90.933	0.727	91.700	0.736	0.185	94.100	0.416	90.600	0.915	0.001
DSSQ TUT	15.167	1.079	16.033	1.074	0.571	13.400	1.118	14.233	1.069	0.592
DSSQ TRI	19.500	1.238	22.033	1.253	0.156	20.167	1.078	21.433	1.245	0.445
Motivation	15.533	1.248	16.033	0.915	0.787	20.033	1.372	22.000	1.419	0.323
Math difficulty	2.367	0.200	2.233	0.184	0.626	1.800	0.200	2.133	0.164	0.203
Stress	2.700	0.210	2.767	0.248	0.838	2.133	0.248	2.367	0.232	0.495
Fatigue	3.800	0.175	2.633	0.212	< 0.001	3.333	0.211	3.700	0.240	0.256
Positive mood	21.600	1.517	22.633	1.100	0.583	20.933	1.663	21.700	1.736	0.751
Negative mood	17.333	1.319	16.267	0.854	0.500	15.500	1.000	16.200	0.967	0.617

Note. OT = proportion of probe-caught on-task thoughts. TUT = proportion of probe-caught task-unrelated thoughts. TRI = proportion of probe-caught task-related interference. Proactive TRI = proportion of probe-caught TRI regarding task approach or strategy. Reactive TRI = proportion of probe-caught TRI regarding evaluation of task performance. Math accuracy = accuracy on the math verification portion of the OSPAN. DSSQ TUT = score on the TUT subscale of the DSSQ (out of 40). DSSQ TRI = score on the TRI subscale of the DSSQ (out of 40). Motivation = score on the Success Motivation subscale of the DSSQ (out of 35). Math difficulty = self-rated perceived difficulty of the math portion of the OSPAN (out of 5). Stress = self-rated perceived stress or anxiety during the OSPAN (out of 5). Fatigue = self-rated perceived fatigue during the OSPAN (out of 5). Positive mood = score on the positive affect scale of the PANAS (out of 50). Negative mood = score on the negative affect scale of the PANAS (out of 50).

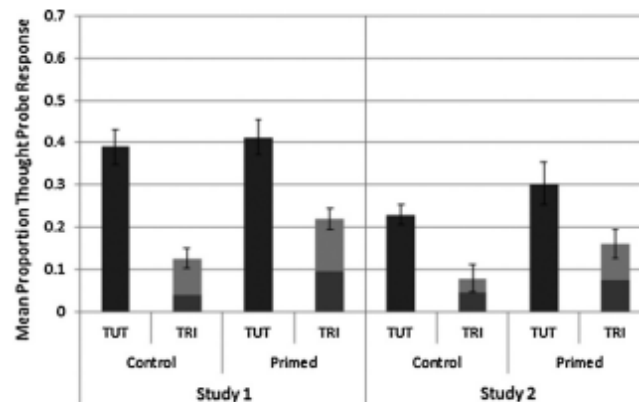


Figure 1. Mean proportions of probe-caught thought types for Study 1 and Study 2 during the OSPAN. TRI is broken down into proactive TRI (thinking about task strategy or approach) which represented by the lighter portions of the TRI columns and reactive TRI (thinking about task performance) which is represented by the darker portion of the TRI columns. SE bars on the TRI columns represent the SE for overall mean proportion of TRI.

We expected that individuals could engage in TRI regarding task strategy (“proactive TRI”) or TRI regarding evaluating one’s task performance (“reactive TRI”). Participants primed for stereotype threat engaged in both of these subtypes of TRI. Primed participants reported a higher proportion of TRI regarding task evaluation ($M = 0.123$, $SE = 0.020$) compared to the control group ($M = 0.087$, $SE = 0.017$), $F(1, 58) = 1.923$, $p = 0.171$, although this difference was not statistically significant. Primed participants also reported a significantly higher proportion of TRI regarding task approach or task strategy ($M = 0.097$, $SE = 0.023$) than control participants ($M = 0.040$, $SE = 0.010$), $F(1, 58) = 5.158$, $p = 0.027$, $d = 0.571$. These results suggest that not only do those under threat potentially reflect more upon their task performance, but they may also think more about ways to improve task performance than those who do not experience threat.

Although we had thought that proactive TRI might be beneficial for task performance, proactive TRI was *not* positively correlated with math verification accuracy in either the control condition ($r = -0.098, p = 0.605$) or the stereotype threat primed condition ($r = 0.085, p = 0.715$).⁴

Although Mrazek et al. (2011) determined that their probe-caught mind-wandering reports were more closely related to retrospective TUTs in participants primed for stereotype threat, we found using online thought probes that stereotype threat primed and control participants did not differ in terms of probe-caught TUTs ($F(1, 58) = 0.160, p = 0.691$). Primed participants were off-task more than control participants, but group differences in mind-wandering were driven by differences in TRI rather than TUTs.

Although we found that stereotype primed participants reported increased TRI and decreased on-task thoughts compared to controls, we did not replicate the standard stereotype threat effect on task performance. Contrary to our initial hypothesis, threat participants ($M = 90.933, SE = 0.727$) did *not* have worse math verification accuracy than control participants, ($M = 91.700, SE = 0.736$), $F(1, 58) = 1.801, p = 0.185$. Letter recall accuracy also did not differ between the two conditions ($F(1, 58) = 0.243, p = 0.624$).

6.2. Post-task survey measures

Stereotype threat priming did not have an effect on the Task Unrelated Thinking ($p = .571$) and Task-Related Interference ($p = 0.156$) subscales of the DSSQ, although primed participants reported significantly more probe-caught TRI. This suggests that participants may not always be able to accurately report mind-wandering after time has elapsed between the mind-wandering experience and the mind-wandering assessment (Smallwood & Schooler, 2006). In both conditions, scores on the TRI subscale of the DSSQ were not correlated with probe-caught TRI ($r = 0.000, p = 0.997$ for controls and $r = -0.063, p = 0.811$ for primed participants).

Although it has been proposed that increased anxiety may be one mechanism by which stereotype threat undermines task performance (Schmader et al., 2008), we found no effect of stereotype threat priming on the post-task question about perceived stress ($F(1, 58) = 0.472, p = 0.838$). However, primed participants did perceive the OSPAN to be more fatiguing than did control participants ($F(1, 58) = 18.006, p < 0.001, d = 1.12$). We found no effect of stereotype threat on self-reported positive affect ($F(1, 58) = 0.305, p = 0.583$) or negative affect ($F(1, 58) = 0.461, p = 0.500$). Likewise, the two conditions did not differ in terms of self-reported motivation ($p = 0.787$) or perceived difficulty of math verification ($p = 0.626$) OSPAN trials.

7. Study 2 goals

Study 1 addressed the impact of priming personal, *performance-related* concerns on mind-wandering and task performance using a stereotype threat manipulation. As predicted, priming stereotype threat resulted in increased mind-wandering, particularly TRI. While participants primed for threat reported more TRI in general, participants in both conditions reported mind-wandering about task evaluation as well as task strategy. Although participants primed for threat reported fewer on-task thoughts than controls, they did not report more TUTs. The stereotype

threat manipulation cued performance-related concerns, but did not increase concerns about things unrelated to the experimental task. This pattern of results is consistent with the “control failures × current concerns” framework of mind-wandering (2010).

We did not find an effect of stereotype threat priming on OSPAN performance in Study 1, contrary to our expectations. Participants primed for threat did as well on the math verification portion of the OSPAN as the control participants. Math verification accuracy and letter recall accuracy were quite high for both groups. It is possible that the mathematical component of the OSPAN, which involved set sizes of 2 or 3 relatively simple math verification problems, was not challenging enough to induce stereotype threat-related performance deficits. Although stereotype threat primed participants reported significantly more TRI, they did not report more stress or negative affect than control participants, which are mechanisms believed to contribute to performance deficits in individuals under threat (Schmader et al., 2008; Schmader et al., 2009).

Study 2 was designed to replicate Study 1 and to assess the impact that stereotype threat has on mind-wandering and performance using a more challenging task. We predict that stereotype threat primed participants will again report more TRI (but not necessarily more TUTs) and will have worse OSPAN math verification accuracy compared to control participants.

8. Study 2 method

8.1. Participants

As in Study 1, sixty female undergraduate students from the University of North Carolina at Greensboro participated in exchange for course credit (M age = 19.25, SD = 1.12). All of these participants were included in the analyses described below. As in Study 1, male participants were also tested in both conditions, but only the data of female participants were included in the following analyses¹.

8.2. Procedure

We primed math-gender stereotype in the same way as Study 1. In Study 2, however, participants completed a more challenging version of the OSPAN that included 81 trials and had set sizes varying between 3, 4, and 5. Nine thought probes appeared at quasi-random intervals during the OSPAN and participants were again instructed to use their keypads to indicate what they were thinking about immediately before the probe appeared. The thoughts probe used were the same as in Study 1 and probes were placed in the same locations within the OSPAN task as in Study 1. As in Study 1, OSPAN trials containing a thought probe were dropped from the analyses reported. The same post-task surveys and questions that were included in Study 1 were included in Study 2.

9. Study 2 results and discussion

9.1. Performance and mind-wandering

Study 2 replicated the mind-wandering findings of Study 1 using a more challenging experimental task. As shown in Table 1 and Figure. 1, probe-caught TRI was higher in the stereotype threat primed participants ($M = 0.161$, $SE = 0.035$) than in the control participants ($M = 0.080$, $SE = 0.032$), $F(1, 58) = 5.527$, $p = 0.022$, $d = 0.423$. Control participants were more likely to indicate being on-task ($M = 0.541$, $SE = 0.043$) than stereotype threat primed participants ($M = 0.352$, $SE = 0.051$), $F(1, 58) = 7.916$, $p = 0.007$, $d = 0.733$. Again, although the stereotype threat participants reported more TRI than controls, the groups did not differ in reported TUTs, $F(1, 58) = 1.826$, $p = 0.182$. The group differences in mean proportion of on-task thoughts were driven by group differences in TRI rather than TUTs.

As in Study 1, participants reported both subtypes of TRI. Although group differences were not statistically significant, primed participants reported numerically more TRI regarding task evaluation ($M = 0.084$, $SE = 0.015$) compared to control participants ($M = 0.035$, $SE = 0.021$), $F(1, 58) = 3.600$, $p = 0.063$. Although this difference did not reach statistical significance it is possible that those under threat reflected more upon their task performance. Although this difference did not reach statistical significance, primed participants also reported numerically more TRI regarding task strategy ($M = 0.077$, $SE = 0.017$) than control participants ($M = 0.045$, $SE = 0.017$), $F(1, 58) = 1.772$, $p = 0.188$. Those under threat may have also been more likely to think about ways to improve task performance. However, engaging in proactive TRI did not seem to improve performance. Probe-caught proactive TRI was not positively correlated with task performance in either condition ($r = -0.268$, $p = 0.153$ for control and $r = -0.096$, $p = 0.613$ for stereotype threat primed) (see Footnote 4).

Using a more challenging OSPAN task, we replicated the standard stereotype threat effect on task performance. Participants primed for math-gender stereotype threat ($M = 90.600$, $SE = 5.014$) had worse math verification accuracy than control participants ($M = 94.100$, $SE = 2.280$), $F(1, 58) = 12.11$, $p = 0.001$, $d = 0.164$. There were no group differences in letter recall accuracy, $F(1, 58) = 0.268$, $p = 0.790$.

9.2. Other measures

As in Study 1, we found no effect of stereotype threat priming on retrospective TUTs and TRI as measured by the DSSQ. Again, participants may not always be able to accurately report instances of mind-wandering after some time has elapsed between the mind-wandering experience and the mind-wandering assessment (Smallwood & Schooler, 2006). As in Study 1, we did not find that scores on the TRI subscale of the DSSQ correlated with probe-caught TRI in the control participants ($r = 0.260$, $p = 0.166$). However, DSSQ TRI scores did correlate with probe-caught TRI in the stereotype threat primed participants ($r = 0.379$, $p = 0.039$). We found no condition differences for the post-task questions regarding perceived stress ($F(1, 58) = 0.472$, $p = 0.495$), perceived fatigue ($F(1, 58) = 1.315$, $p = 0.256$), positive affect ($F(1, 58) = 0.102$, $p = 0.751$), and negative affect ($F(1, 58) = 0.253$, $p = 0.617$). Likewise, the two conditions did not differ in terms of reported motivation to do the task well (as measured by the Motivation subscale of the DSSQ; $p = 323$) and perceived difficulty of the math verification portion of the OSPAN ($p = 0.203$).

10. General discussion

The present studies demonstrate that priming of personal, performance-related current concerns can increase mind-wandering, particularly mind-wandering regarding task strategy and task evaluation (TRI). In Study 1, female undergraduates primed for math-gender stereotype threat reported more TRI than female undergraduate controls. In Study 2, which employed a more challenging task, female undergraduates primed for math-gender stereotype threat reported more TRI than female undergraduate controls and also had worse performance on the mathematical component of the experimental task. As noted in Footnote 1, our stereotype threat manipulation appears to have differentially affected male and female participants. For both studies, male participants tested in the control and stereotype threat conditions did not differ from each other in terms on TRI or math verification performance, while female participants in the two conditions did. This suggests that the experimental manipulation did not induce general, math-related anxiety in participants that were tested in the control condition, but rather specifically primed performance-related concerns in female participants under stereotype threat.

In addition to supporting the “control failures \times current concerns” framework of mind-wandering (McVay & Kane, 2010), these studies support current theories of stereotype threat. The Study 2 findings are consistent with an explanation that stereotype threat undermines task performance through increases in metacognitive thoughts, which can result in the depletion of executive control resources (Blascovich, Spencer, Quinn, & Steele, 2001; Croizet et al., 2004; Inzlicht, McKay, & Aronson, 2006; Schamder & Johns, 2003). While past work has suggested that stereotype threat may decouple one’s attention from the current task in a way that leads to increases in task-unrelated thoughts (Mrazek et al., 2011), the present studies find that stereotype threat increases thoughts related to the current task.

In the current studies, participants primed for stereotype threat did not report increased stress and increased negative mood compared to controls, which are proposed mechanisms by which stereotype threat leads to increased off-task thoughts and worse performance on the stereotyped task. In addition to work in the stereotype threat literature suggesting that increases in anxiety and suppression of worry-laden thoughts impair task performance, work within the mind-wandering literature also suggests that valence predicts whether off-task thoughts will disrupt performance. Benny and Banks (2015) demonstrated that negatively valenced TRI episodes in particular predict poor performance, at least on higher demand tasks. In the current studies, we did not have participants rate the emotional valence of their mind-wandering episodes. It is also worth noting that in the current studies, it remains unclear whether our stereotype type threat manipulation influences participants’ thought content by specifically increasing worry-laden thoughts or by increasing task diligence in those under threat. Future studies can be done to further investigate how emotion interacts with mind-wandering to affect task performance (particularly in those under stereotype threat) and whether stereotype threat interventions increase TRI through prompting distracting worry-laden thoughts or by more generally increasing task vigilance.

The thought probe methodology used in the current studies offers a more nuanced way to assess mind-wandering than methodologies used in past studies of stereotype threat and intrusive thoughts. Memory-demanding retrospective thought content questionnaires and verbal report coding schemes carry the risk of mind-wandering episodes being overlooked or forgotten by participants. Using more detailed online thought probes, we lend support to previous research

demonstrating that stereotype threat specifically leads to increased *task-related* worries and metacognitions (Beilock et al., 2007; Cadinu et al., 2005). We did find that not only did stereotype threat participants report intrusive thoughts about task performance evaluation, but they also reported more intrusive thoughts about *task strategy*. Although we found no evidence in the current studies that TRI about task strategy aided performance, it is possible that mind-wandering about task strategy may have downstream benefits for task performance on different tasks, longer tasks, or across multiple task sessions. Future studies can be done to determine under which situations mind-wandering about task strategy can benefit task performance.

We found differences in the breakdown of proactive and reactive TRI. Stereotype threat primed participants in Study 1, which had a lower task demand, reported numerically more proactive TRI than control participants. Stereotype threat primed participants in Study 2, which had a higher task demand, reported numerically more reactive TRI. It is possible that, at least in threatened individuals, tasks with a higher demand leads to more evaluative of task performance, but not more thinking about task strategies. More work can be done to determine how threat and task demands differentially alter these two different types of TRI.

We labeled TRI about task strategy as “proactive TRI” because we believed that participants could think about strategies that they could implement on upcoming trials of the experimental task. Although participants reported thinking about task strategy, it is possible that they were not actually thinking about potentially beneficial strategies to implement on future trials. Instead, participants could have been evaluating past or current strategies when they endorsed proactive TRI on the thought probes. We did not necessarily expect evaluative thoughts about task strategy to lead to improved task performance within the current task. Participants may adopt either effective or ineffective strategies when completing the OSPAN (Dunlosky & Kane, 2007). It is possible that even if participants in these studies thought about new strategies to use on future OSPAN trials, they were not generating effective strategies that would improve performance. More research is needed to determine the exact content of participants’ thoughts when they endorse TRI on online thought probes.

Overall, the current studies provide a test with replication of the “control failures x current concerns” framework of mind-wandering proposed by McVay and Kane (2010). In addition to demonstrating that priming personal, performance-related concerns in younger adults leads to increased TRI, the current studies can provide insight into why older adults consistently report less overall mind-wandering but more TRI than younger adults (Frank et al., 2015; Giambra, 1989; Grodsky & Giambra, 1990; Jackson & Balota, 2012; Krawietz et al., 2012; McVay & Kane, 2012a; McVay et al., 2013). The results of the current younger adult studies suggest that stereotype threat increases TRI. Future research can examine whether memory-related stereotype threat may account for the increased TRI typically reported by older adults. The “control failures x current concerns” framework (McVay & Kane, 2010) can be further tested by systemically priming *everyday concerns* in younger and older adults, and examining whether this manipulation increases TUT reports.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.concog.2017.08.002>.

REFERENCES

- Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *British Journal of Psychology*, 59(4), 423–430. <http://dx.doi.org/10.1111/j.2044-8295.1968.tb01157.x>.
- 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(2), 256–276. <http://dx.doi.org/10.1037/0096-3445.136.2.256>.
- Benny, B., & Banks, J. B. (2015). Under pressure: An examination of the predictors of choking. *Journal of Individual Differences*, 36, 93–100. <http://dx.doi.org/10.1027/1614-001/a000160>.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225–229. <http://dx.doi.org/10.1111/1467-9280.00340>.
- 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.x>.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786. <http://dx.doi.org/10.3758/BF03196772>.
- 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 & Social Psychology Bulletin*, 30(6), 721–731. <http://dx.doi.org/10.1177/0146167204263961>.
- Dixon, R. A., & Hultsch, D. F. (1984). The metamemory in adulthood (MIA) instrument. *Psychological Documents*, 14, 3.
- Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, 60(9), 1227–1245.
- Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, 30, 266–278. <http://dx.doi.org/10.1037/pag0000031>.
- Giambra, L. M. (1989). Task-unrelated thought frequency as a function of age: A laboratory study. *Psychology and Aging*, 4(2), 136–143. <http://dx.doi.org/10.1037/0882-7974.4.2.136>.

- Gonçalves, O. F., Rêgo, G., Oliveira-Silva, P., Leite, J., Carvalho, S., Fregni, F., ... Boggio, P. S. (2017). Mind wandering and the attention network system. *Acta Psychologica, 172*, 49–54. <http://dx.doi.org/10.1016/j.actpsy.2016.11.008>.
- Grodsky, A., & Giambra, L. M. (1990–91). The consistency across vigilance and reading tasks of individual differences in the occurrence of task-unrelated and task-related images and thoughts. *Imagination, Cognition and Personality, 10*(1), 39–52. <http://dx.doi.org/10.2190/6QG5-CXVV-4XUR-7P3K>.
- Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science, 17*(3), 262–269. <http://dx.doi.org/10.1111/j.1467-9280.2006.01695.x>.
- Jackson, J. D., & Balota, D. A. (2012). Mind wandering in younger and older adults: Converging evidence from the sustained attention to response task and reading for comprehension. *Psychology and Aging, 27*(1), 106–119. <http://dx.doi.org/10.1037/a0023933>.
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*(7), 614–621. <http://dx.doi.org/10.1111/j.1467-9280.2007.01948.x>.
- Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological Science*. <http://dx.doi.org/10.1177/0956797617706086>.
- Klinger, E., & Cox, W. M. (1987). Dimensions of thought flow in everyday life. *Imagination, Cognition, and Personality, 7*(2), 105–128. <http://dx.doi.org/10.2190/7K24G343-MTQW-115V>.
- Krawietz, S. A., Tamplin, A. K., & Radvansky, G. A. (2012). Age and mind wandering during text comprehension. *Psychology and Aging, 27*(4), 951–958. <http://dx.doi.org/10.1037/a0028831>.
- Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The persistence of thought: Evidence for a role of working memory in the maintenance of task-unrelated thinking. *Psychological Science, 23*, 375–380. <http://dx.doi.org/10.1177/0956797611431465>.
- Matthews, G., Joyner, L., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three? In I. Merville, I. J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 335–350). Tilburg: Tilburg University Press.
- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 196–204. <http://dx.doi.org/10.1037/a0014104>.

- McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2005) and Watkins (2008). *Psychological Bulletin*, *136*(2), 188–197. <http://dx.doi.org/10.1037/a0018298>.
- McVay, J. C., & Kane, M. J. (2012a). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*(2), 302–320. <http://dx.doi.org/10.1037/a0025250> (Advance online publication).
- McVay, J. C., & Kane, M. J. (2012b). Drifting from slow to “D'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive- control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 525–549. <http://dx.doi.org/10.1037/a0025896>.
- McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind- wandering across controlled and ecological contexts. *Psychonomic Bulletin & Review*, *16*(5), 857–863. <http://dx.doi.org/10.3758/PBR.16.5.857>.
- McVay, J. C., Meier, M. E., Touron, D. R., & Kane, M. J. (2013). Aging ebbs the flow of thought: Adult age differences in mind wandering, executive control, and self- evaluation. *Acta Psychologica*, *142*(1), 136–147. <http://dx.doi.org/10.1016/j.actpsy.2012.11.006>.
- 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*(6), 1243–1248. <http://dx.doi.org/10.1016/j.jesp.2011.05.011>.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*(1), 14–29. <http://dx.doi.org/10.1027/1618-3169.54.1.14>.
- Schmader, T., Forbes, C. E., Zhang, S., & Mendes, W. B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality and Social Psychology Bulletin*, *35*(5), 584–596. <http://dx.doi.org/10.1177/0146167208330450>.
- 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](http://dx.doi.org/10.1037/0033-295X.115.2.336).
- Schuster, C., Martiny, S. E., & Schmader, T. (2015). Distracted by the unthought – Suppression and reappraisal of mind wandering under stereotype threat. *PLoS ONE*, *10*(3), <http://dx.doi.org/10.1371/journal.pone.0122207>.
- Seli, P., Carriere, J. S. A., Levene, M., & Smilek, D. (2013). How few and far between? Examining the effects of probe rate on self-reported mind wandering. *Frontiers in Psychology*, *4*, 430. <http://dx.doi.org/10.3389/fpsyg.2013.00430>.

- Smallwood, J. M., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., Oconnor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*(4), 657–690. <http://dx.doi.org/10.1016/j.concog.2004.06.003>.
- Smallwood, J. M., Fitzgerald, A., Miles, L. K., & Phillips, L. H. (2009). Shifting moods, wandering minds: Negative moods leads the mind to wander. *Emotion*, *9*(2), 271–276. <http://dx.doi.org/10.1037/a0014855>.
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144–1150. <http://dx.doi.org/10.3758/MC.36.6.1144>.
- Smallwood, J., O'Connor, R. C., & Heim, D. (2005). Rumination, dysphoria, and subjective experience. *Imagination, Cognition and Personality*, *24*(4), 355–367.
- Smallwood, J. M., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. <http://dx.doi.org/10.1037/0033-2909.132.6.946>.
- Steele, C. M., & Aronson, J. (1995). Contending with a stereotype: African-American intellectual test performance and stereotype threat. *Journal of Personality and Social Psychology*, *69*(5), 797–811. <http://dx.doi.org/10.1037/0022-3514.69.5.797>.
- Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., & Baddeley, A. (1995). Stimulus-independent thought depends on central executive resources. *Memory and Cognition*, *23*(5), 551–559. <http://dx.doi.org/10.3758/BF03197257>.
- Watson, D., Clark, L. A., & Tellegan, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

¹ Although we were primarily interested in the effect of math-gender stereotype threat on TRI and task performance, data from male participants was also collected for both studies and for both conditions. For Study 1 we collected data from 49 male participants (24 control and 25 experimental). For Study 1, a 2x2 factorial ANOVA for probe-caught TRI with gender and condition included as between-subjects factors revealed a gender by group interaction ($F(1,112) = 9.202, p = 0.003$). However, there was no main effect of gender ($p = 0.083$), although female participants ($M = 0.17, SD = 0.13$) reported numerically more TRI than male participants ($M = 0.13, SD = 0.10$). There was also no main effect of group ($p = 0.456$). If the stereotype threat manipulation inadvertently lead to general math anxiety in participants within the experimental group, we might have expected a group difference, with experimental participants reporting more TRI than control participants. This was not the case. Control participants ($M = 0.14, SD = 0.12$) reported roughly the same amount of TRI as participants tested in the stereotype threat condition ($M = 0.17, SD = 0.12$).

A 2×2 factorial ANOVA for math verification accuracy with gender and condition included as between-subjects factors did not reveal a main effect of group ($p = 0.921$). Control participants ($M = 91.54, SD = 3.70$) and experimental participants ($M = 91.40, SD = 3.62$) had very similar math verification performance. Again, if the stereotype threat manipulation inadvertently lead to general math anxiety in participants within the experimental group, we might have expected a group difference, with experimental participants performing more poorly on the math verification portion of the OSPAN than control participants. The factorial ANOVA likewise did not reveal a main effect of gender ($p = 0.642$). Female ($M = 91.32, SD = 3.99$) and male ($M = 91.65, SD = 3.19$) participants had similar math verification performance. There was no group * gender interaction ($p = 0.970$).

Male participants in the study 1 control condition ($M = 0.15, SD = 0.10$) and experimental condition ($M = 0.10, SD = 0.10$) did not differ significantly from each other in terms of probe-caught TRI ($p = 0.505$), suggesting that the stereotype threat manipulation did not lead to math anxiety or an increase in performance-related concerns in male participants within the experimental group that received a stereotype threat manipulation. In the study 1 control condition, male participants ($M = 0.16, SD = 0.10$) and female participants ($M = 0.13, SD = 0.13$) did not differ significantly in terms of probe-caught TRI ($p = 0.076$). This suggests that female control participants were not experiencing a greater degree of performance-related concerns than male participants.

Study 2 included data from 35 male participants (18 control and 17 experimental). A 2×2 factorial ANOVA for probe-caught TRI with gender and condition as between-subjects factors revealed a gender by group interaction ($F(1,95) = 5.600, p = 0.020$) but not main effects of gender ($p = 0.928$) or group ($p = 0.532$). Female participants ($M = 0.13, SD = 0.17$) and male participants ($M = 0.13, SD = 0.14$) reported the same amount of probe-caught TRI. Control participants ($M = 0.11, SD = 0.11$) reported numerically less TRI compared to participants tested in the stereotype threat condition ($M = 0.15, SD = 0.19$), but this difference was not significant.

A 2×2 factorial ANOVA for math verification accuracy with gender and condition included as between-subjects factors revealed a gender by group interaction ($F(1,95) = 5.583, p = 0.020$) and a main effect of group ($F(1,95) = 6.016, p = 0.016$), with control participants ($M = 93.65, SD = 2.33$) obtaining higher math verification accuracy than experimental participants ($M = 91.40, SD = 3.65$). There was no main effect of gender ($p = 0.488$). Female ($M = 92.35, SD = 4.25$) and male ($M = 92.86, SD = 2.32$) participants had similar math verification performance. The fact that control female participants performed similarly on the math task to control male participants suggests that female participants within the control group were not experiencing stereotype threat.

Males in the control ($M = 0.16, SD = 0.10$) condition reported more TRI than males in the experimental condition ($M = 0.11, SD = 0.16$), but this difference was not significant ($p = 0.717$). Males in the control ($M = 92.89, SD = 2.30$) and experimental ($M = 92.82, SD = 2.40$) conditions likewise did not differ from each other in terms of OSPAN math verification accuracy ($p = 0.878$). Again, this would suggest that the stereotype threat manipulation did not lead to increased math anxiety in males within the experimental condition. Control condition males ($M = 92.89, SD = 2.30$) and females ($M = 94.10, SD = 2.28$) did not differ in terms of OSPAN math verification accuracy ($p = 0.965$). Again, the fact that control female participants performed similarly on the math task to control male participants suggests that female participants within the control group were not experiencing stereotype threat. Control condition males ($M = 0.16, SD = 0.10$) and females ($M = 0.08, SD = 0.10$) also did not differ significantly in terms of probe-caught TRI ($p = 0.973$).

The patterns above support the experimental manipulation as impacting TRI frequency during the task due to gender-based stereotype threat regarding math performance.

² In addition to the measures described within the text of this article, the following additional measures were included. These are not central to the study but are reported here in the interest of full and open reporting. The Interest subscale of the Dundee Stress State Questionnaire (DSSQ; Matthews, Joyner, Gililand, Huggins, & Falconer) was included as a measure of task interest, along with a single-item post-task question about task interest. In addition to a single-item question about perceived difficulty of the math verification portion of the OSPAN, participants responded to a single-item question about their perceived overall OSPAN task difficulty. Participants answered a single-item question about how focused they were on accurately verifying math equations and how focused they were on accurately recalling letters during the OSPAN. To further measure task effort and motivation, participants responded to single-item post-task questions about how satisfied they were with their OSPAN performance, how much effort they put into the OSPAN, and how well they thought they did on the OSPAN.

Participants also responded to a single-item post-task question regarding how distracted they felt that they were during the OSPAN, whether they had to read the OSPAN task instructions multiple times before understanding them, and whether they believed a negative stereotype regarding females being inferior in mathematical abilities existed in society. Finally, participants responded to the Locus, Anxiety, and Achievement scales of the Metamemory in Adulthood Questionnaire (MIA; Dixon & Hultsch). In both studies, the means for these variables did not differ between the control and primed conditions (all p values > 0.05).

³ We predicted that participants primed for math-gender stereotype threat would be slower to pair the categories of “female” and “math” together than they would to pair the categories of “male” and “math” together. We found that primed participants in Study 1 were slower to pair the categories of “female” and “math” together compared to control participants in Study 1 ($F(1, 58) = 7.374, p = 0.009$). In Study 2, the group differences were marginal ($F(1, 58) = 3.637, p = 0.064$), but primed participants were likewise slower to pair the categories of “female” and “math” together. These IAT findings combined with the mind-wandering differences obtained in both studies indicate that our priming manipulation successfully elevated math-gender stereotype threat in our stereotype threat condition.

⁴ To further examine the effect of different thought types on OSPAN performance, we compared math verification and letter recall accuracy on OSPAN sets immediately following probes where proactive TRI was reported, probes where reactive TRI was reported, and probes where TUTs were reported. These analyses suggest that engaging in proactive TRI is not more beneficial for immediate task performance than engaging in reactive TRI or TUTs.

One-way ANCOVAs were conducted to determine the effect of thought probe response (proactive TRI, reactive TRI, and TUTs) on performance for sets immediately following probes where mind-wandering was reported, controlling for set size. For OSPAN sets following a thought probe where mind-wandering was reported, two measures of accuracy were calculated. A measure of math verification accuracy was calculated as the number of correctly verified math problems out of the total number of math problems within a set. For each set, participants got the letter recall portion correct if they were able to recall all letters within the set in correct serial order.

For Study 1, there was no significant effect of thought type on set-level math verification accuracy when controlling for set size, $F(3, 330) = 0.451, p = 0.637$. Math verification accuracy on sets following probes where proactive TRI was reported ($M = 0.9356, SD = 0.9356$) was no different than accuracy on sets following probes where reactive TRI ($M = 0.9281, SD = 0.2074$) and TUTs ($M = 0.9100, SD = 0.1825$) were reported. Similarly, there was no significant effect of thought type on set-level letter recall accuracy when controlling for set size, $F(3, 330) = 0.447, p = 0.640$. Letter recall accuracy on sets following probes where proactive TRI was reported ($M = 0.9167, SD = 0.2803$) was no different than accuracy on sets following probes where reactive TRI ($M = 0.9403, SD = 0.2387$) and TUTs ($M = 0.9031, SD = 0.2965$) were reported.

The same pattern of results was observed for Study 2. There was no significant effect of thought type on set-level math verification accuracy when controlling for set size, $F(3, 330) = 0.313, p = 0.731$. Math verification accuracy on sets following probes where proactive TRI was reported ($M = 0.9413, SD = 0.1377$) was no different than accuracy on sets following probes where reactive TRI ($M = 0.9279, SD = 0.1541$) and TUTs ($M = 0.9148, SD = 0.1936$) were reported. Similarly, there was no significant effect of thought type on set-level letter recall accuracy when controlling for set size, $F(3, 330) = 0.248, p = 0.781$. Letter recall accuracy on sets following probes where proactive TRI was reported ($M = 0.9286, SD = 0.2598$) was no different than accuracy on sets following probes where reactive TRI ($M = 0.9245, SD = 0.2667$) and TUTs ($M = 0.9441, SD = 0.2306$) were reported.