CLARIFYING MOTIVATION IN STEREOTYPE THREAT: THE CHRONIC THREAT MODEL

By

Anthony Trevor Misisco

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Master of Arts

2021

ABSTRACT

CLARIFYING MOTIVATION IN STEREOTYPE THREAT:
THE CHRONIC THREAT MODEL

By

Anthony Trevor Misisco

Stereotype threat (ST) has become a well-documented phenomenon (Spencer, Steele, & Quinn, 1999; Steele, 1997; Steele & Aronson, 1995), but its specific mechanisms remain unclear. It is commonly accepted that ST harms task performance through a cognitive process pathway wherein working memory capacity is impaired (Schmader & Beilock, 2012; Schmader, Johns, & Forbes, 2008). However, there is a lack of consensus regarding the motivational pathway of ST, as some researchers suggest that ST motivates individuals to dispel stereotypes (e.g., Schmader et al., 2008) and others suggest that ST is demotivating (e.g., Walton, Murphy, & Ryan, 2015). This research attempted to reconcile these conflicting motivational accounts, suggesting that motivation is initially bolstered when ST is experienced and eventually undermined under conditions of chronic ST. Drawing on Lindsley, Brass, and Thomas's (1995) conceptualization of efficacy-performance spirals, a chronic process model of stereotype threat is presented. Individuals participated in a multiple-trial goal-regulation task in which individuals set and pursue goals, adjusting goals after each trial in response to performance feedback. Partial support was found for several components of the proposed model. The present study contributes to the literature by (a) integrating disparate portrayals of motivation within ST research; (b) delineating ST processes within individuals and across time, utilizing a process-oriented model; and (c) examining ST within the scope of goal-setting, an organizationally-relevant mediating mechanism through which ST may impair performance in real-world settings.

ACKNOWLEDGEMENTS

I am gracious for my friends and family that have helped keep me sane through the challenges of completing my thesis. I appreciate my friends at Michigan State, Sergio and Sarah, for their support as I concluded my time at MSU. I am grateful for the new friends and support systems I have developed after my transition to the University of South Florida, especially Alyssa; they all made my move in the midst of a pandemic survivable. Friends I have made in happenstance, such as Jensine at a SIOP poster session, are connections I couldn't do without. Finally, I would like to thank my thesis chair, Steve W.J. Kozlowski, and my committee members, Kevin Ford and Daisy Chang, for their flexibility and assistance as I navigated challenges unique to completing my thesis. Many people offered instrumental and emotional support along the way, and I feel motivated to succeed and ensure their support was well-placed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Stereotype threat (ST) occurs when an individual is at risk of confirming a negative group stereotype (Steele & Aronson, 1995). Widely known negative stereotypes exist for several social groups. For example, women are stereotyped as less competent in mathematics (Spencer et al., 1999) and white men are viewed as unathletic (Stone, Perry, & Darley, 1997). The threat of being judged for performing consistently with ingroup cultural stereotypes is anxiety-provoking; above and beyond facing potential stereotypical treatment from others, the individual may also fear the prospect of damaging their self-concept in the event of poor performance, gaining threatening information about their ability in the domain of the stereotype.

Stereotype threat has been conceptualized as activating "propositional relations" (Gawronski & Bodenhausen, 2006) between an individual's ingroup, ability domain, and self-concept (Schmader et al., 2008). Threat typically primes an individual's group membership, and concurrently primes stereotypes about their group's domain competence. Thus, individuals that wish to perform well and have a positive self-concept in the domain face a disconnect: they believe they are like other members of their group, but they believe they have an ability that their group stereotypically does not. ST is important from an organizational standpoint because even fairly innocuous conditions – which likely bear more resemblance to conditions that would exist within the workplace – have been found to trigger ST. Illustratively, Danahar and Crandall (2008) found that when women simply marked their gender after a math test rather than before taking it, the gender performance gap shrunk by a third.

Most stereotype threat research identifies cognitive impairments as the key mediators of ST on performance deficits (for reviews, see Casad & Merritt, 2014; Schmader & Beilock, 2012; Schmader, Johns, & Forbes, 2008). Specifically, working memory capacity is impaired due to

the tendency for those experiencing ST to experience a stress response, monitor their environment for threatening cues, and regulate their affective reactions (Schmader et al., 2008). Working memory is the system in which information is temporarily stored and manipulated during cognitive tasks (Baddeley, 1992). It is conceptualized as a limited, resource-based system; it cannot attend to unlimited stimuli. ST causes individuals to grapple with the prospect of fulfilling negative group stereotypes, resulting in fewer resources being available to accomplish the performance task.

Although identifying the cognitive pathway of ST performance deficits has been an important development, particularly for experimental research, it does not address perhaps the most compelling reason to study stereotype threat. Stereotype threat is a proposed mechanism through which several societal inequalities surface. Steele (1997) suggested that ST may explain gender differences in the pursuit of STEM education as well as gender- and race-based differences in academic performance, as consistent long-term ST may result in disidentification and withdrawal from the domain. Longitudinal studies frequently make this association. For example, Woodcock, Hernandez, Estrada, and Schultz (2012) found that across three academic years, minority students who experienced stereotype threat were more likely to disidentify with the scientific domain and consequently less likely to pursue a scientific career. This body of literature has a notable trend that contrasts it from laboratory work: ST is explained in terms of motivational constructs, not cognitive impairments. The cognitive perspective of stereotype threat may explain short-term evaluative threat, but it appears that motivational factors may be the prominent drivers of longitudinal ST effects.

Further review of the ST literature reveals an unresolved disconnect. Stereotype threat appears capable of either increasing or decreasing motivation. ST frequently appears to cause

individuals to work harder (Grand, 2017; Jamieson & Harkins, 2007; Oswald & Harvey, 2000; Schmader & Beilock, 2012; Schmader et al., 2008; Seitchik & Harkins, 2015), suggesting they are motivated to dispel the negative group stereotype (commonly referred to as "stereotype reactance"). For example, Oswald and Harvey (2000) unexpectedly found that threatened individuals attempted more problems during a mathematics test than unthreatened individuals. Similarly, threatened individuals in Grand's (2017) study processed more "contacts" in a radar tracking simulation. While the previous paragraph identified that withdrawal is often associated with ST (indicating reduced motivation), there is also evidence that ST can bolster motivation. Consequently, this has prompted a hunt for potential moderators explaining when stereotype threat causes reactance and when it results in withdrawal.

I suggest that time is the neglected factor. Temporal processes are infrequently conceptualized in organizational research despite the richness of temporally-specified models (Sonnentag, 2012). Indeed, organizational scholars are rarely interested in static phenomena; rather, research and theory are meant to capture longitudinal processes (Ployhart & Vandenberg, 2010). Reviewing the literature, it appears that variance in motivation may be attributable to the differing timeframes in which ST is measured. Generally, it appears that stereotype reactance most often occurs in static laboratory designs (see Table 1). Presented with a short-term, novel task, it is plausible that stereotype threat will motivate individuals to work harder. However, exposure to long-term chronic stereotype threat is likely demotivating. Those under stereotype threat are more prone to both actively monitoring the task environment for failure-centered information as well as becoming preoccupied with affective reactions to the task, using these affective reactions as indicators of failure (Amodio et al., 2004; Johns et al., 2008). Mangels et al. (2012) also found electrophysiological evidence that threatened individuals react more

strongly to negative feedback. In operational settings, individuals may be constantly vigilant of threatening information; when they make mistakes or fail to achieve goals, they may be more likely to interpret their failure as diagnostic of their ability and subsequently reduce effort. Steele, Spencer, and Aronson (2002) suggested that disengagement could lead to chronic disidentification, resulting in a cyclical disconnect from the domain. Cross-sectional and single-trial research may fail to capture the dynamic, time-sensitive experience of stereotype threat.

I define *acute stereotype threat* as a temporally bound, single instance of experienced stereotype threat. Kalokerinos et al. (2014) observed that most laboratory studies of stereotype threat assess the effect of acute stereotype threat. In a prototypical stereotype threat experiment, participants are made aware of an existing negative stereotype about their group, often an explicit remark about their group's ability in the target domain. When they perform the experimental task, they typically suffer a performance deficit as a result. While relevant to instances such as testing scenarios, acute stereotype threat may poorly approximate regular, repetitive experiences of threat in organizational life. I define *chronic stereotype threat* as frequent, recurrent instances of experienced stereotype threat. In an organizational setting, this can manifest in a variety of ways. If an individual's stereotyped group is underrepresented, this could serve as a persistent reminder that their group isn't valued in the domain (Saenz, 1994). The physical environment, such as office décor, may also provide identity-relevant cues (Walton et al., 2015). This definition does not require that the threat cues are persistently available in the environment. If an individual experiences an acute, blatant threat targeting their ability on a performance task, this threat may still be psychologically available across repeated instances of task performance.

In this research, I contribute to the literature by attempting to resolve the discrepant motivational findings in the stereotype threat literature. A model articulating the discrete experiences of acute and chronic stereotype threat is proposed that elucidates when threat experiences are expected to be acute or chronic in nature. Importantly, the model moves beyond typical approaches to studying stereotype threat that recite conditions necessary for the phenomenon to occur. Such strategies overemphasize predicting when ST will occur at the expense of understanding how and why it occurs. Instead, I describe a within-person process through which stereotype threat is expected to manifest into motivational and performance deficits. Lastly, this model is evaluated with a laboratory study to provide a preliminary test of its propositions.

The body of this paper is organized as follows. First, a high-level review of the broad stereotype threat literature will be provided, followed by a brief review of the cognitively-oriented ST literature and a more focused review of the extant motivation-oriented ST literature. Next, I will argue that traditional applications of stereotype threat towards understanding organizational phenomena are insufficient for understanding how ST may affect the day-to-day life of individuals. Finally, a chronic model of stereotype threat will be proposed and evaluated with empirical data, integrating the observation in short-term "acute" threat conditions that motivation is boosted with the tendency for long-term "chronic" threat conditions to undermine motivation.

Stereotype Threat

Stereotype threat refers to a situation in which an individual feels at risk of confirming, as self-characteristic, a negative stereotype about one's group (Steele & Aronson, 1995). Steele and Aronson (1995) elucidated this possibility with a plight faced by African American students. When these students perform in an academic scenario, they risk "…confirming or being judged by a negative societal stereotype — a suspicion — about their group's intellectual ability and competence" (p.797). Several hundred laboratory-based studies have shown that stereotype threat can cause performance deficits (Nguyen & Ryan, 2008; see also Ryan & Nguyen, 2017; cf. Zigerell, 2017). Most frequently, cognitive ability tests are the domain explored and the stereotypes involve women and ethnic minorities.

Steele (1997, p. 617-618) identifies several general features of stereotype threat. First, stereotype threat can occur when there exists a widely known negative stereotype about one's group and the individual is aware of this stereotype. Southern-accented individuals in the United States are frequently stereotyped as less intelligent than Northerners (Kinzler & DeJesus, 2013); this knowledge can cause stereotype threat to occur if the individual fears their behavior will confirm the stereotype. Threat can occur even if the individual does not endorse the stereotype (Steele, 1997). Fearing that others will construe their behavior as stereotype-consistent is sufficient to feel threatened. Next, if the task or situation that one is engaged in makes the stereotype salient, it is more likely that individuals will perceive the context as evaluative. Steele (1997) used classroom demographic composition as an example. When stereotyped and nonstereotyped individuals are both present in an evaluative setting, it may make the stereotype especially salient. Relatedly, underrepresentation has been frequently tied to threat reactions (e.g. Avery, 2003; Murphy, Steele, & Gross, 2007). When stereotyped individuals are poorly

represented in a domain, they may perceive this as a signal that their group does not have the ability to perform in the domain and that they do not belong (Alderfer, 1987). Additionally, this "token" status can cause individuals to feel that their performance will be viewed by nonstereotyped members as indicative of their group's (in)ability, presenting additional concerns that may distract them from performing well (Saenz, 1994). Although mixed representation in an evaluative context is not necessary nor sufficient for stereotype threat to occur (Steele, 1997), it is frequently used to invoke stereotype threat in laboratory settings and appears to be a real concern for organizations (Walton et al., 2015).

Along with both the recognition that a negative stereotype exists and sufficient situational strength of this stereotype, Steele (1997) identifies two other interrelated conditions that affect the occurrence of stereotype threat: the individual must be identified with the performance domain, and the task must be sufficiently difficult to challenge the individual's ability. The thesis of Steele's (1997) seminal work is that (a) academic identification is crucial for school success; (b) societal pressures such as poor access to quality education, traditional gender roles, and a dearth of resources restricts academic identification development for several societal groups; and (c) stereotyped group members who overcome these hurdles and become domain identified face the additional prospect of stereotype threat. Therefore, Steele expected that only those individuals who are highly domain identified will experience stereotype threat. Individuals who are not identified will not experience threat because performance in the domain is inconsequential for their self-definition. Task difficulty is particularly important because highly identified, well-achieving members of the stereotyped group are confident in their ability — their past experience suggests that they should do well on the task. If the task difficulty is at the upper bound of the individual's ability, performance frustration signals a "newly met limit to their

ability" (p. 620). Failure in this context could cause the threatened individual to accept the stereotype as self-evident. This may be especially frustrating since the individual is heavily invested in the domain.

There is general support for these two tenets. In a meta-analysis of stereotype threat effects on cognitive ability test performance, Nguyen and Ryan (2008) found that low math-identified women experienced the smallest deficit in performance from stereotype threat (d = .11). However, moderately-identified women unexpectedly experienced the greatest deficit in performance (d = .52), counter to the hypothesis that highly-identified women (d = .29) would suffer the most from stereotype threat. Nguyen and Ryan (2008) suggested that this departure from Steele's (1997) theoretical work could be due to stereotype threat sparking reactance among highly-identified women, who would be especially motivated to dispel the stereotype. Alternatively, inconsistent operationalization of domain identification was also suspected of producing this discrepancy. As expected, test difficulty also moderated performance, such that the threat effect was larger when tests were more difficult. This finding was consistent with individual studies examining test difficulty as a moderator (e.g. O'Brien & Crandall, 2003).

Cognitive Mechanisms of Stereotype Threat

Schmader et al. (2008) provide the dominant process model of ST (see also Beilock & Schmader, 2012; Casad & Merritt, 2014). The model includes cognitive, affective, and physiological pathways, and also provides some discussion of motivational processes, but the model assigns primacy to cognition in ST. In their model, working memory capacity is the proximal antecedent of impaired performance outcomes in cognitive and social tasks. Working memory is conceptualized as a resource-limited system (Baddeley, 1992). Therefore,

performance impairments in ST are observed when (a) the task is sufficiently difficult to require all of an individual's attentional resources and (b) ST draws some of these resources away from the task domain.

Three interrelated mechanisms are proposed to act on working memory: a physiological stress response, vigilance, and suppression processes (Casad & Merritt, 2014; Schmader et al., 2008). Physiological activation is commonly observed in ST (e.g. Ben-Zeev, Fein, & Inzlicht, 2005; O'Brien & Crandall, 2003) and is commonly associated with performance deficits, especially when tasks are difficult (e.g., Bernstein-Bercovitz, 2003). ST also causes individuals to monitor their performance, as the imbalanced propositional relationships between their self-, group-, and domain-concepts elicit excessive vigilance to disambiguating information (Schmader et al., 2008). This leads to more deliberate, controlled information processing (Seibt & Förster, 2004) and increased sensitivity to failure-related cues. ST is likely to lead to monitoring for information that will help the individual decide whether they are behaving stereotypically, resulting in self-oriented attention that may tax working memory capacity and impair performance (Schmader et al., 2008; see also Kluger & DeNisi, 1996). Finally, ST causes individuals to actively suppress negative thoughts and feelings that result from monitoring and appraisal processes. Suppression is suggested to be an effortful, taxing process (Muraven & Baumeister, 2000), further detracting from working memory resources. Taken together, these processes appear to act on working memory capacity, resulting in diminished performance outcomes.

Motivation in Stereotype Threat Research

Schmader et al. (2008) suggest that stereotype threat motivates individuals to disconfirm a negative group stereotype when it is directed toward them. According to Schmader et al. (2008), "…activating negative stereotypes about a social identity one possesses motivates individuals to try to combat the stereotype but … this creates some sort of extra situational burden that interferes with the ability to perform as well at a task as might otherwise be possible" (p. 337-338). This was proposed based on findings from a few lines of research. Several studies have found that threatened individuals surprisingly tend to perform better than controls when tasks are easier and more readily learned (e.g., Ben-Zeev et al., 2005; O'Brien & Crandall, 2003). This has been interpreted as evidence that ST is motivating. If a task is not sufficiently difficult to tax working memory, the motivational surplus will outweigh any cognitive deficit associated with task performance. Thus, the typical performance impairment found in ST research appears to be related to reductions in working memory capacity, not due to a lack of motivation. Forbes, Schmader, and Allen (2008) further articulated this process. They found that threatened individuals who identified with their academic domain were especially vigilant in monitoring for performance errors and were subsequently able to respond to errors more efficiently, providing further evidence that motivation is frequently boosted under threat conditions, and can even facilitate performance in some circumstances. See Table 1 for a list of studies assessing motivation in laboratory setting.

While motivation is somewhat tertiary to the model presented by Schmader et al. (2008), Jamieson, Harkins, and colleagues are the dominant proponents of a motivational account of stereotype threat. Based on Harkin's (2006) original work, the two have built a substantial research stream exploring their "mere effort" account of stereotype threat (Brown & Harkins,

2016; Huber, Seitchik, Brown, Sternad, & Harkins, 2015; Jamieson & Harkins, 2007, 2009, 2011; McFall, Jamieson, & Harkins, 2009; Seitchik & Harkins, 2015; Seitchik, Jamieson, & Harkins, 2012). According to the mere effort account, threatened individuals perceive that they will be evaluated based on their task performance, motivating them to perform well. This causes individuals to utilize a "prepotent" response strategy. Individuals adopt the most likely, dominant, obvious task strategy available. When the prepotent response is the most appropriate strategy, performance will be facilitated. However, adopting a prepotent task strategy can also cause individuals to fail to recognize advantageous strategies, resulting in diminished performance.

To summarize, lab-based motivation research in the ST literature tends to suggest that counterintuitively, threatened individuals often undermine performance through their efforts to disconfirm negative group stereotypes (Jamieson & Harkins, 2009, 2011; Seibt & Förster, 2004). They tend to work "harder" but not "smarter" (Grand, 2017), adopting prepotent response strategies that may lead to immediate success at the expense of discovering more effective strategies. Threatened individuals appear capable of modifying their prepotent responses when they are able to recognize the response is incorrect (McFall et al., 2009), but there is also evidence that threat can cause inflexibility in task strategy utilization (Carr & Steele, 2009). Taken together, this body of research advances the notion that stereotype threat bolsters motivation, and consequent performance outcomes are dependent on the adaptiveness of this mobilization.

Contrasting the previously reviewed findings, the experience of threat over time is suggested to cause domain disidentification and withdrawal (Steele, 1997; Steele & Aronson, 1995). Based on this rationale, Steele (1997) advanced the notion that stereotype threat could

Table 1: Key Laboratory-based Studies of Motivation in the Stereotype Threat Literature

Reference	Operationalization	Motivation Increased?
Forbes, Schmader, & Allen (2008)	Neuronal indices of error monitoring on the Eriksen-flankers task.	Yes
Grand (2017)	Number of "contacts" processed in a radar tracking simulation.	Yes
Harkins (2006)	Adoption of prepotent response strategy on the Remote Associates Task.	Yes
Huber, Seitchik, Brown, Sternad, & Harkins (2015)	Adoption of prepotent response strategy on a rhythmic ball bouncing task.	Yes
Hirnstein, Freund, & Hausmann (2015)	Verbal fluency performance.	Yes
Jamieson & Harkins (2007)	Adoption of prepotent response strategy on the antisaccade task.	Yes
Jamieson & Harkins (2009)	Adoption of prepotent response strategy on GRE quantitative problems.	Yes
Jamieson & Harkins (2011)	Adoption of prepotent response strategy on the Stroop task.	Yes
Jamieson & Harkins (2012)	Adoption of prepotent response strategy on GRE quantitative problems.	Yes
Kray, Thompson, & Galinsky (2001)	Performance in a negotiation task and negotiation goal level.	Yes; when ST was explicit, not implicit
Latsch & Hannover (2014)	Self-reported learning goal orientation.	Yes
Mangels, Good, Whiteman, Maniscalco, & Dweck (2012)	Number of clicks made in online learning module.	No; decreased
McFall, Jamieson, & Harkins (2009)	Adoption of prepotent response strategy on the Stroop task, an anagram task, and the antisaccade task.	Yes
O'Brien & Crandall (2003)	Performance on easy versus difficult mathematics tests.	Yes
Oswald & Harvey (2000)	Number of problems attempted on a mathematics test.	Yes
Seibt & Förster (2004)	Self-reported "general motivation."	Mixed; series of studies reported either no difference or increased motivation
Seitchik & Harkins (2015)	Adoption of prepotent response strategy on mental arithmetic tasks.	Yes
Seitchik, Jamieson, & Harkins (2014)	Adoption of prepotent response strategy on GRE quantitative problems.	Yes
Stone (2002)	Handicapping behavior in athletes.	No; decreased

partially explain social group differences in society. In stark opposition to the lab-based studies, which show frequent motivation increases, I was unable to identify a single study displaying improved motivation in a longitudinal design.

Stereotype threat has been related to racial differences in intentions to pursue a science-related career (Woodcock, Hernandez, Estrada, & Schultz, 2012), as well as gender differences in science career interest (Deemer, Thoman, Chase, & Smith, 2014), procrastination in STEM classes (Deemer, Smith, Carroll, & Carpenter, 2014), and interest in a computer science task (Smith, Sansone, & White, 2007). Delisle, Guay, Senécal, and Larose (2009) found in an eighteen-month longitudinal design that when women are enrolled in science programs with low female representation, they are more likely to experience stereotype threat. Additionally, when women reported experiences of chronic stereotype threat, they also reported lower confidence in their ability to achieve career goals (von Hippel, Issa, Ma, & Stokes, 2011). Finally, Osborne and Walker (2006) found in a longitudinal study of inner-city high school students that academic identification was differentially predictive of withdrawal for Caucasian and African American students. While academic identification was associated with reduced likelihood of withdrawal for Caucasian students, the reverse was true for African American students. It was suggested that this occurred due to experiencing stereotype threat, although stereotype threat was not measured. This appears consistent with Steele's (1997) prediction that stereotype threat is most damaging for domain-identified, threatened students. In sum, studies using both (a) longitudinal designs and (b) cross-sectional designs in operational settings where chronic threat conditions are present suggest that chronic threat exposure tends to be associated with diminished motivation.

Stereotype Threat in Organizational Research

Within the field of industrial/organizational psychology, study of stereotyping and discrimination has focused primarily on inequities produced by organizational decision-makers, such as in selection and assessment settings. These discussions have often provoked great debate, especially surrounding the generalizability of lab studies (both the techniques used and the real-world validity) as well as how readily the small effects occasionally found may snowball into substantial outcomes (Grand, Golubovich, Ryan, & Schmitt, 2013; Landy, 2008; Martell, Lane, & Emrich, 1996; Sackett & Ryan, 2012; Steele & Davies, 2003). Stereotype threat, specifically, has received somewhat less attention within IOP. A primary research area explores the potential for ST to influence the selection and testing settings (Steele & Davies, 2003; Walton et al., 2015). A significant proportion of traditional stereotype threat research investigates performance in test situations, often utilizing standardized tests in laboratory experiments (Steele, 2010), so examining testing within a selection context was a natural extension of the literature. Real-world employment-testing contexts have received comparatively less attention, and concerns still exist regarding the generalizability of lab findings to operational settings (Sackett & Ryan, 2012; Shewach, Sackett, & Quint, 2018).

Even if generalizability concerns are resolved for test settings, the extant stereotype threat research poorly approximates the effects of ST during everyday organizational life. Stereotype threat effects are thought to occur in a recursive, cyclical manner: performance-feedback cycles cause further performance decrements, feedback oversensitivity, and task disengagement (Cohen & Garcia, 2008). These decrements are proposed to occur longitudinally as individuals are subject to chronic stereotype threat (Nussbaum & Steele, 2007). Despite this assumption, very little empirical evidence exists to support this process. Most stereotype threat research is single-

trial experimental research or cross-sectional survey research. Process cannot be measured in these designs, only hypothesized. Although single-instance occurrences of ST may affect individuals in settings such as selection tests, ST during everyday operations should manifest differently. ST in an organizational setting is persistent; it may be built into the organization, such as through minority representation (Walton et al., 2015). Persistent ST should accumulate and have downstream effects on those experiencing it.

In their focal article in *Industrial and Organizational Psychology*, Kalokerinos, von Hippel, and Zacher (2014) similarly make a distinction between one-time "acute" experiences of ST and longer-lasting "chronic" threat. They suggest that some researchers have prematurely concluded that stereotype threat effects are small or null based on studies of acute threat, failing to recognize the cumulative nature of chronic threat. They concluded that the debate swirling around stereotype threat has dissuaded researchers from exploring real-world, organizationally-relevant consequences of ST.

Establishing the recursive nature of ST is especially important for articulating the process by which individuals "deindividuate" from a domain. ST is commonly proposed to drive societal inequities through its demotivating effect: individuals perform more poorly in a domain, withdraw from it, and stunt their development (Cohen & Garcia, 2008). Yet, experimental studies typically suggest an opposite effect: individuals experiencing stereotype threat often work harder during an experimental task to dispel the negative group stereotype, although they typically perform more poorly.

Although these separate lines of research appear to offer contradictory motivational perspectives, they can be reconciled. Individuals may initially be motivated to dispel stereotypes

in novel[1] task environments. Conversely, it appears that chronic, long-term exposure to stereotypes will demotivate individuals as ST encourages domain disidentification. Even in short-lived experimental settings, reduced interest (a motivation-like construct) has been found by the end of the study (Davies, Spencer, Quinn, & Gerhardstein, 2002; Smith, Sansone, & White, 2007). Nussbaum and Steele (2007) found that threatened individuals were able to maintain motivation by (counterintuitively) disengaging from the situation during a single-trial task, which allowed them to preserve their self-esteem and discount threatening feedback. They suggested that disengagement may be adaptive in the short-run but maladaptive under chronic conditions.

Most ST studies do not take a process-oriented approach; they consist of a single performance episode and the process mechanisms are hypothesized rather than measured. There is a dearth of lab-based research examining the process mechanisms through which ST impairs performance and diminishes motivation in performance domains across trials and across time. That is, few studies assess ST in a dynamic context (see Grand, 2017 for an exception). Examining changes in motivation requires repeated-measures, within-person designs (Ployhart & Vandenberg, 2010), which receive little attention in ST research. As Samuelson, Fernandez, and Grand's (2017) article title astutely notes, “Life doesn’t happen at the between-person level.” Consequently, motivation has been overlooked as a mediator of performance within ST research.

[1] I use the word “novel” to refer to a task situation in which a person has not yet directly been exposed to stereotype threat. The task situation cannot be truly unique, as no existing stereotype would apply to performance in that domain. However, if the individual has already experienced chronic threat in the domain, they may have already disengaged. For example, we would describe a stock-prediction task as a novel task environment: it is unlikely that the participants have performed this specific stock-prediction task, but performance will likely be affected by broader stereotypes about mathematic and quantitative reasoning.

MODEL OF CHRONIC STEREOTYPE THREAT

I now turn to building a model of chronic stereotype threat. Lindsley et al.'s (1995) model of efficacy-performance spirals provides a useful framework for understanding how cyclical relationships may exist between variables and across time. According to this model, there is a positive, cyclical relationship between self-efficacy and performance. Because of the interdependence between these variables, they will tend to move downward or upward together in a "spiral" fashion. In a downward spiral, a poor performance episode leads to a reduction in self-efficacy, which further reduces performance and subsequent self-efficacy in the next performance cycle. Lindsley et al. (1995) suggest that across time, these spirals can result in dramatic swings in performance.

Lindsley et al. (1995) identified factors that should influence the probability of the occurrence and continuation of spirals. Several are relevant to stereotype threat conditions. First, Lindsley et al. (1995) suggest that accurate, timely, specific feedback will allow actors to revise task-related strategies and adjust their level of effort, thus avoiding the occurrence of spirals. In the case of threatened individuals, they tend to be especially frustrated by feedback (Schmader et al., 2008), which has been tied to both inflexible strategy usage (Carr & Steele, 2009) and reduced effort (Mangels, Good, Whiteman, Maniscalco, & Dweck, 2012; see also *Hypothesis 3*). Lindsley et al. (1995) also hypothesized that spirals are more likely to continue if individuals develop internal, stable, and uncontrollable performance attributions. Stereotype threat causes individuals to grapple with the prospect of fulfilling group stereotypes; failure can cause individuals to accept these stereotypes as self-evident (Schmader et al., 2008; see also *Hypotheses 4* and *5*). Automatic information processing is also expected to relate to the continuation of spirals. Individuals who use automatic information processing use superficial,

disengaged performance strategies rather than reflecting on the task. In addition to the feedback-related evidence that this will be more likely to occur for threatened individuals, recall that Jamieson, Harkins, and colleagues (e.g., Jamieson & Harkins 2007, 2009) found that threatened individuals tend to use prepotent response strategies, relying on automatic rather than thoughtful processing. Finally, Lindsley et al. (1995) postulated a positive relationship between spiral continuation and emotional arousal. Significant evidence exists relating stereotype threat to emotional arousal (e.g. Amodio et al., 2004; Johns et al., 2008). In summary, there appears to be substantial reason to suspect that threatened individuals will suffer downward, recursive, cyclical performance effects.

The chronic stereotype threat model is presented in Figure 1. My review of Lindsley et al.'s (1995) efficacy-performance spiral model offers a framework supporting the hypothesis that the cyclical, recursive effects theorized by some researchers (Cohen & Garcia, 2008; Kalokerinos et al., 2014) will surface. The chronic stereotype threat model distinguishes between initial threat conditions ("acute" stereotype threat) and chronic threat conditions. The model can be summarized as follows. Initially, individuals experiencing acute stereotype threat will be especially motivated to dispel their negative group stereotype. However, their motivation will quickly drop-off in response to performance feedback, especially when this feedback is negative. They will be less effective at incorporating feedback into future task strategies, further undermining performance and subsequent motivation. When individuals experience chronic stereotype threat, they will set lower performance goals, put forth less effort towards accomplishing their goals, achieve lower levels of performance, receive more negative feedback and interpret this feedback more negatively, attribute failure to stable, internal characteristics, and experience reduced self-efficacy. The model is cyclical, such that the downward deviations

in these variables affect the next performance cycle. Each component of the model will now be reviewed.

Following stereotype threat induction, it is expected that individuals experiencing stereotype threat will set higher performance goals and exhibit higher effort. This is consistent with the previously reviewed static laboratory studies, which often find higher levels of initial motivation following ST (Forbes et al., 2008; Jamieson & Harkins, 2007, 2009, 2011). However, considering the summary above, exposure to chronic stereotype threat should cause threatened individuals to withdraw more quickly from the task, setting lower performance goals and exhibiting lower effort across time. See Appendix B1 for a graphical summary.

Motivation is a broad, umbrella "meta-construct" (Kanfer et al., 2017), encompassing concepts such as initiation, direction, intensity, and persistence (Campbell & Pritchard, 1976; Pinder, 1984). I used self-set goal level, effort, and goal commitment as an operationalization of motivation, consistent with other researchers (e.g., Ilies & Judge, 2005) I view the current study's treatment of motivation as stronger than many operationalizations within the stereotype threat literature (see Table 1). Many ST studies rely on perceptual measures of motivation, and other "trace" and/or outcome measures such as number of problems attempted, number of task targets processed, etc. Furthermore, my method is consistent with Kanfer et al's (2017) conceptualization of modern motivation theory: "Three features distinguish this contemporary theory from older views, such as (1) primacy of goals, (2) emphasis on goal pursuit and associated affective processes, and (3) conception of motivation as an active process…" Thus, utilizing goal-centered measures of motivation appears consistent with modern motivation theory.

Hypothesis 1: Individuals experiencing *acute* stereotype threat will initially (a) set higher performance goals; (b) report higher levels of effort than those in the control condition; and (c) report higher goal commitment than those in the control condition.

Hypothesis 2: Individuals experiencing *chronic* stereotype threat will (a) set lower performance goals; (b) report lower levels of effort across time than those in the control condition; and (c) report lower goal commitment than those in the control condition.

Despite the increase in motivation that often accompanies stereotype threat, it is typically found that stereotype threat reduces performance due to the cognitive process pathway reviewed previously (so long as the task is sufficiently complex, O'Brien & Crandall, 2003). This pattern of results has been described as "working harder," but not "working smarter" (Grand, 2017). However, the relationship between ST and performance is often moderate in single-trial, static designs (Shewach et al., 2018), which approximates the pattern of results we expect under initial conditions.

Hypothesis 3: Individuals experiencing acute stereotype threat will initially demonstrate poorer task performance than those in the control condition.

Feedback is an important component of goal-setting (Locke & Latham, 2002). People require information regarding their goal progress to reflect on their current state and adjust their performance strategies. However, previous research has shown that individuals experiencing stereotype threat engage in suboptimal feedback seeking strategies (Roberson, Deitch, Brief, & Block, 2003) and frequently discount feedback, especially if they believe that feedback is related to prejudice against their social group (Crocker & Major, 1989; Major, Spencer, Schmader, Wolfe, & Crocker, 1998). Making errors or failing to progress towards a goal at a satisfactory

pace can be anxiety provoking (Bell & Kozlowski, 2008; Kanfer & Heggestad, 1999; Keith & Frese, 2005; Locke & Latham, 2002). When individuals experience ST, they may find it especially difficult to approach feedback constructively rather than experience negative affective and motivational outcomes (Grand, 2017; Keller & Dauenheimer, 2003; Krendl, Richeson, Kelley, & Heatherton, 2008; Wraga et al., 2007). Threatened individuals tend to be drawn towards negative stimuli (Forbes et al., 2008; Schmader et al., 2008), suggesting that feedback may be especially frustrating for threatened individuals when they are not making adequate progress towards performance goals. Those under stereotype threat are more prone to both actively monitoring the task environment for failure-centered information as well as becoming preoccupied with affective reactions to the task, using these affective reactions as indicators of failure (Amodio et al., 2004; Johns et al., 2008). Mangels et al. (2012) also found electrophysiological evidence that threatened individuals react more strongly to negative feedback. These individuals had difficulty regulating their emotions, and heightened arousal ultimately resulted in task disengagement.

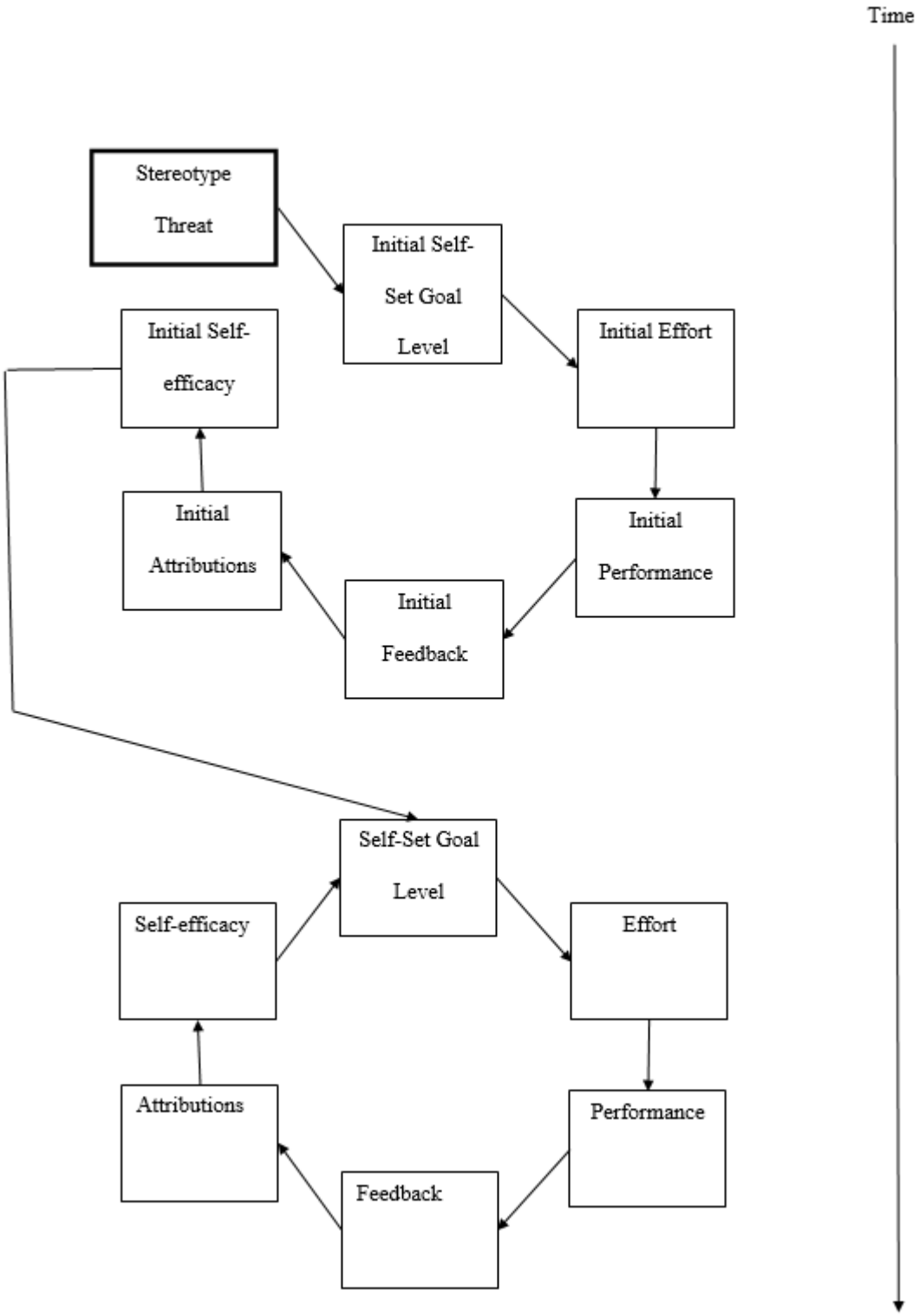
Those experiencing stereotype threat may find it especially difficult to approach feedback constructively rather than experience negative affective and motivational outcomes (Grand, 2017; Keller & Dauenheimer, 2003; Krendl, Richeson, Kelley, & Heatherton, 2008; Wraga et al., 2007). Threatened individuals tend to be drawn towards negative stimuli (Forbes et al., 2008; Schmader et al., 2008), suggesting that feedback may be especially frustrating for threatened individuals when they are not making adequate progress towards performance goals. Those under stereotype threat are more prone to both actively monitoring the task environment for failure-centered information as well as becoming preoccupied with affective reactions to the task, using these affective reactions as indicators of failure (Amodio et al., 2004; Johns et al.,

2008). Mangels et al. (2012) also found electrophysiological evidence that threatened individuals react more strongly to negative feedback. These individuals had difficulty regulating their emotions, and heightened arousal ultimately resulted in task disengagement.

This mirrors a research stream in the feedback/performance literature. In a meta-analysis, Kluger and DeNisi (1996) found that, although feedback interventions generally improve performance by about one-half of a standard deviation, over one-third of interventions decreased performance. They proposed that feedback can undermine performance if it directs attention away from the task and towards the self. It is likely that this will happen more frequently to individuals experiencing stereotype threat as they attempt to resolve the imbalance provoked by ST (e.g., I identify with my group, and I want to perform well, but there is a stereotype that my group does not perform well at this task) (Schmader et al., 2008). In summary, threatened individuals are more likely to both receive more negative feedback and to generally react more negatively to feedback.

Hypothesis 4: Individuals experiencing chronic stereotype threat are more likely than controls to (a) react to feedback cues in an especially negative fashion and (b) rate feedback cues as less useful.

Figure 1: Process Model of Chronic Stereotype Threat



Likely related to these feedback perceptions are the subsequent attributions made about the feedback. Performance attribution research offers support for this hypothesis (Martinko, Harvey, & Dasborough, 2011). Those under ST may be initially motivated to dispel negative stereotypes if they believe that they are capable of doing so. However, ST creates a state of cognitive imbalance as individuals engage with dissonant self- and group-stereotypes (Schmader et al., 2008). When individuals under ST receive negative feedback, especially under chronic conditions, they may come to accept the stereotype as self-relevant (especially if they were engaging in increased effort), demotivating and demoralizing them. Indeed, Steele (1997, p. 620) suggested a similar mechanism in his seminal work. Steele described stereotype threat as establishing an "interpretive frame;" if threatened individuals experience performance frustration, they may acknowledge the group-based stereotype as indicative of their own ability. Importantly, this signaling can occur even for individuals who are highly identified with the performance domain, as performance frustration presents a "possibly newly met limit to their ability." They may come to view their performance as uncontrollable and/or stable (a consequence of their group identity) and withhold effort. Others have written about causal attributions within the goal-setting literature (Donovan & Williams, 2003; Ilgen & Davis, 2000; Ilies & Judge, 2005). The general conceptual linkage is that after an individual fails to achieve a performance goal, they will be more likely to maintain their effort rather than decrease their goals if they believe that nonattainment was due to unstable, controllable, situational conditions. If they attribute nonattainment to stable, uncontrollable, internal causes, they will be less likely to maintain effort.

Hypothesis 5: Individuals experiencing chronic stereotype threat are more likely than controls to attribute goal nonattainment to stable, internal, uncontrollable causes.

Those that experience stereotype threat often internalize stereotypes, resulting in diminished self-efficacy (Steele, 1997). Bandura and Cervone (1986) found that self-efficacy is vital following receiving negative feedback, as self-efficacy will predict subsequent goal-setting difficulty. Triggering identities associated with negative stereotypes is hypothesized to impair performance expectations and confidence, diminishing performance (Baumeister, Hamilton, & Tice, 1985; Cheryan & Bodenhausen, 2000). A primary causal mechanism at play is likely motivational, as impaired expectations would lead to withdrawal and reduced task-directed effort (Kanfer, 1990; Steel & König, 2006; Vroom, 1964). Some studies have found that this self-efficacy reduction is related to reduced performance (Desrichard & Köpetz, 2005), whereas others have not found a diminishing effect (Schweinle & Mims, 2009; Spencer, Steel, & Quinn, 1999). I believe that self-efficacy will be particularly important in a goal-setting situation for the reasons described next.

There has been some contention surrounding self-efficacy and performance outcomes (Bandura, 1977, 1986, 1989, 1991, 1997, 2012; Bandura & Locke, 2003; Carver & Scheier, 1998; Locke & Latham, 1990; Powers, 1973, 1991; Vancouver & Purl, 2017; Vancouver et al., 2001; Vancouver, Thompson, Tischner, & Putka, 2002). The primary argument employed amongst those skeptical of self-efficacy's role in positively affecting performance is the potential for overconfidence to develop, leading to a reduction in effort and resources towards the task at hand. Indeed, Lindsley et al. (1995) note that positive, upward performance spirals could lead to overconfidence and complacency. However, while these arguments center around the results of *increases* in self-efficacy, self-efficacy drops below baseline after experiencing ST. Whereas higher self-efficacy may be associated with overconfidence, drops in self-efficacy more likely incite frustration, performance anxiety, and poor task strategy development (Locke & Latham,

2002). Furthermore, both Bandura's Social-Cognitive Theory (1986, 2012) and Control Theory (Carver & Scheier, 1998; Vancouver and colleagues, 2001, 2002, 2017) appear to accept that downward deviations in self-efficacy are associated with impaired performance outcomes. Although some researchers have not found a drop in self-efficacy following ST manipulations (e.g., Fogliati & Bussey, 2013; Spencer et al., 1999), I believe that the goal-setting domain will be likely to provoke efficacy-relevant thoughts, especially if participants continually receive negative performance feedback.

Hypothesis 6: Individuals experiencing chronic stereotype threat are more likely than controls to experience reduced self-efficacy.

Finally, it is expected that the cumulative result of these hypothesized mechanisms will be diminished downstream performance for individuals experiencing ST.

Hypothesis 7: Individuals experiencing chronic stereotype threat will perform more poorly than controls across time.

METHOD

Individuals participated in a stock prediction task explained below in the Fall of 2018 and Spring 2019. To invoke stereotype threat conditions, 303 females were the participant sample for this study. Widely held stereotypes in the United States suggest that women are less competent in mathematical, quantitative domains than men (Spencer et al., 1999), so stereotype threat laboratory studies frequently use quantitative tasks with female participants (Steele, 1997). Participants were recruited from the MSU SONA research pool. Participants were randomly placed in an experimental (Stereotype Threat) or a control condition.

Task

Participants completed the multiple cue probability learning (MCPL) task popularized by Earley et al. (1989). This task is frequently used in motivation research (e.g. Beck & Schmidt, 2012; Fisher & Ford, 1998; Park, Schmidt, Scheu, & Deshon, 2007). In this task, individuals predict the value of company stock prices. Individuals are presented with three pieces of information to make their decisions: the company's marketing performance, research and development performance, and production division performance. The company's real value is determined by an underlying linear regression equation of which the participants are unaware; the same equation is used to determine the value of all stocks during the session. For example, Earley et al. (1989) used the equation (which I also utilized):

$$Y_e = 0X_1 + .33X_2 + .67X_3 + \text{error}$$

Where X_1 represented the company's marketing performance, X_2 represented research and development, and X_3 represented the production division. The information provided was in percentage form relative to each department's goals. For example, a company's marketing department may have performed at the 75% goal level, their research and development department performed at 105% of their goal level, and the production division performed at 90%. The error term was randomly drawn, with a mean of 0 and a range of -$10 to $10. Participants were told that company stock prices ranged between $10 and $150.

Procedure

Individuals were seated in a classroom in mixed-gender[2] sessions of 10-20 individuals. This design choice was made to maximize the amount of data collected per session and, more importantly, to create an evaluative context that will make the gender stereotype especially salient. Upon arrival, participants were greeted and provided with a Qualtrics link to access the task on their personal computers. Participants received an introduction to the task and were instructed to raise their hands if they had any questions. Consistent with Earley et al. (1989), participants were informed that they should not base their stock estimates on any past experiences or prior knowledge. Rather, they should approach the task without making any assumptions. The entire task was programmed to display through Qualtrics.

The experiment took place over 126 trials. Trials were grouped into 21 blocks of six trials each. Before each block, participants set a goal by estimating how far away their stock estimate would be from the real value. Participants had 10 seconds from the time the cue was displayed to enter an estimate for the stock's value. After each block, participants received authentic performance feedback displaying the distance between their estimates and the actual stock values for each of the six trials within the block.

Manipulation

The stereotype threat manipulation was adapted from past research (Grand, 2017; Jamieson & Harkins, 2007; Spencer et al., 1999). In the stereotype threat condition, participants were instructed that the purpose of the study was to examine why women perform more poorly

[2] Women are the target group for analysis, so men's data were not used. Consistent with other studies (e.g., Grand, 2017), men were included to increase the salience of the gender stereotype. The MSU SONA pool trends female, so this resulted in minimal discarded data and improved the task environment's authenticity.

in quantitative and mathematical domains. Further, they were informed that this task has shown gender differences in performance, and those who lack quantitative ability will struggle on the task. To maintain the salience of the stereotype (i.e., to provide chronic ST conditions), individuals in the threat condition were asked to enter their gender before each block for "data quality purposes" (Seibt & Förster, 2004). Individuals in the control condition were not told that there are gender differences on this task. They did not indicate their gender until the end of the study.

Measures: Control Variables

Demographics. Individuals were asked to provide their gender, age, ACT/SAT score, year in college, major, and race. These measures were collected at the end of the experiment to avoid invoking ST in the control group (Danahar & Crandall, 2008).

Cognitive ability. Participants reported their ACT or SAT scores during initial demographic reporting. Although the SAT has been found to overpredict women's college performance (Ramist, Lewis, & McCamley-Jenkins, 1994; Steele, 1997), this is not viewed as an issue in the current study because women are the participants in both conditions. Moreover, the measure is being used to tap general cognitive ability, for which these scores are viewed as valid (Schmidt, 1988). Gully, Payne, Kiechel, and Whitman (1999) found a strong correspondence between actual and reported test scores ($r = .94$).

Trait goal orientation. Trait goal orientation was assessed at the beginning of the study using the 5-item performance-approach measure and 4-item performance-avoid measure developed and validated by VandeWalle (1997). An alternative account of ST suggests that ST effects occur due to their influence on regulatory foci (Grimm, Marman, Maddox, & Baldwin,

2009; Seibt & Förster, 2004). Thus, trait- and state-goal orientation were controlled. Most of this research establishes an approach/avoid dichotomy and does not consider mastery orientation, so mastery orientation was not measured.

State goal orientation. The 5-item performance-prove and performance-avoid orientation scales from Horvath, Scheu, & DeShon (2001) were used and administered intermittently (see Table 2): Before the 1st, 6th, 11th, 16th, and 20th blocks.

Goal commitment. Klein, Cooper, and Monahan (2013) suggest that self-report is typically the best means to measure goal commitment. Before each block, commitment was measured with the 4-item KUT (Klein et al., Unidimensional, Target-free) measure validated in Klein, Cooper, Molloy, and Swanson (2014). Goal commitment is a standard inclusion in goal-setting studies.

State affect. State affect was measured intermittently using a shortened state PANAS measure, following the strategy used by Koy and Yeo (2008) and Yeo, Frederiks, Kiewitz, and Neal (2014). The three items with the highest factor loadings from Watson, Clark, and Tellegen (1988) on both the positive affect and negative affect scales were selected for use, excluding affective states irrelevant to a lab task (e.g., scared).

Trait affect. The 10-item short form version of the Positive and Negative Affect Schedule (PANAS) developed and validated by Thompson (2007) was administered at the beginning of the study.

Domain identification. According to Steele (1997), domain identification is a necessary prerequisite for ST effects to occur. The 9-item mathematic subscale of the domain identification measure developed by Smith and White (2001) was used. Similar to the concern that reporting

demographic information could affect performance, the items in this measure (see Appendix A7) may also cause threat conditions, which would be problematic for examining the differences between the ST and control groups. For this reasoning, the scale was completed at the conclusion of the experiment.

Manipulation check. At the end of the experiment, participants were asked the extent to which they believe that there are gender differences on this task. A 5-item measure was adapted from past research (Bergeron, Block, & Echtenkamp, 2006; Grand, 2017; Steele & Aronson, 1995). Some items from the original scales were not adapted because either they did not have a natural analog for the current study (e.g., "At work, people of my gender often face biased evaluations from others") or because the item appeared to be directed too broadly for the specific context of an experiment ("In [English] classes people of my [race] often face biased evaluations").

Measures: Dependent Variables

Goal level. Individuals set a performance goal prior to each block. Because my hypotheses are time-sensitive, it will be important to give participants a frame of reference for initial goal-setting to ensure that goals are indicative of effort rather than uncertainty. Consistent with Earley et al.'s (1989) "specific, moderate goal" condition, participants were informed that setting a goal to make predictions within $16 of the actual stock values is a reasonable goal. This will be framed as suggestive rather than assigned. Only 15% of participants in a pilot by Earley et al. (1989) were able to achieve a goal to make predictions within $10 by the end of the study after 100 trials. I chose a goal that would be out-of-reach for most participants initially, but would also be achievable and viewed as realistic.

Effort. A 1-item measure of self-reported effort was adapted from Yeo and Neal (2004, 2008). Single item measures are frequently used to measure effort (see also, Maslyn & Uhl-Bien, 2001). Immediately after each block and before any feedback was provided, individuals answered the question "How hard were you trying during the last block?" Participants responded using an 11-point scale ranging from 0 (*not at all*) to 10 (*extremely hard*).

Feedback. Hypothesis 4 stated: "Individuals experiencing chronic stereotype threat are more likely than controls to (a) react to feedback cues in an especially negative fashion and (b) rate feedback cues as less useful. Hypothesis 4A was assessed using the 13-item Positive (7-item) and Negative (6-item) Reaction scales from Brett and Atwater (2001), administered after each block immediately after receiving feedback. Hypothesis 4B was assessed using their responses to a 2-item Feedback Usefulness scale adapted from Brett and Atwater (2001) administered after receiving feedback every block.

Performance Attributions. Attributions were assessed after each block using McAuley, Duncan, and Russell's (1992) 12-item Revised Causal Dimension Scale, as used in past goal-setting research (e.g., Tolli & Schmidt, 2008). The scale assesses the subdimensions of locus of causality, external control, stability, and personal control.

Self-efficacy. Before each block, participants completed a 10-item scale consistent with past recommendations and research (Beck & Schmidt, 2012; Bandura, 1997; Locke & Latham, 1990; Masuda, Locke, & Williams, 2014). This scale asks participants to assess the percent likelihood (from 1 = 10% or less to 10 = 100%) that they will perform at each of 10 performance levels. The performance levels are displayed in increments of $5, such that the participants are asked the percent likelihood that they will estimate the stock price within $50, $45, etc., all the

way to within $5 of the actual stock prices.. Self-efficacy is calculated by averaging these rating scores.

Postexperimental questionnaire. Items from Eagley et al. (1989) were included to assess task strategy usage and participant's use of the different information cues to make their predictions. They were also asked open-ended questions about their self-regulatory strategies and feedback reactions.

Table 2: Summary of Measures and Time of Measurement.

Initial Measures	**Before Blocks (in order of presentation)**	**After Blocks (in order of presentation)**	**Intermittent Measures (1st, 6th, 11th, 16th, 20th blocks)**	**Final Measures**
9-item trait approach + trait avoid	10-item self-efficacy	1-item effort	6-item state affect	Demographics
10-item trait affect	Self-set goal level	13-item Feedback reactions and 2-item usefulness	10-item state approach + state avoid	5- item manipulation Check
	4-item goal commitment	12-item performance attributions		Postexperimental questionnaire
				9-item domain identification

ANALYSIS

The hypotheses explored in this study are time-sensitive. The experimental group was expected to score higher than the control group in early blocks of the experiment on several of the measures of interest, while the control group was expected to score higher closer to the conclusion of the experiment. First, a manipulation check presented at the conclusion of the experiment was analyzed to determine if the manipulation successfully induced Stereotype Threat. Next, omnibus mixed-effect ANOVAs were performed on the focal variables to determine if sufficient variability existed by condition prior to performing more time-sensitive analyses. ANCOVAs were also performed to determine whether this variability still existed after controlling for relevant control variables.

After these initial examinations, t-tests were performed to examine whether the experimental and control groups differed on the focal variables within each block. Participants completed a total of 126 trials arranged within 21 blocks. Eight variables were of focal interest in this study (Effort, Goal level, Performance, Attributions, Feedback Reactions, Feedback Usefulness, Goal Commitment, Self-efficacy), so the results are numerous. To bound the analysis, blocks were organized into four sets of blocks (Blocks 1-5 = Set 1; Blocks 6-10 = Set 2; Blocks 11-15 = Set 3; Blocks 16-21 = Set 4). T-tests were conducted both at the block and the set level, and plots for both sets of analyses are provided in the appendices. However, to (1) simplify reporting, (2) avoid overinterpreting spurious t-tests given the number of planned comparisons, and (3) provide a framework that helps classify blocks as “early” or “late” in the experiment, this organization was adopted and Set-level results are provided in the results section. T-tests were useful to identify whether fine-grained differences existed at the Block and Set levels rather than identify an aggregate measure of group differences.

Next, the lavaan package in R was used to generate latent growth curves. The purposes of this approach were twofold: first, t-tests will tend to underestimate the standard error and inflate the likelihood of identifying a relationship as significant when observations are dependent. Accounting for random effects in growth curve modeling allowed an evaluation of the standard errors relative to those in the t-test analyses and to determine whether they were underestimated previously or whether they provide acceptable estimates. Second, growth curves were estimated that regressed the slope and intercept terms on the Condition variable (i.e., experimental vs. control). While the t-tests indicated whether differences existed at the Block and Set levels, which was crucial due to the time-sensitive nature of my hypotheses, the growth-curve approach provides supplemental aggregate information regarding the conditions across the experiment.

RESULTS

Before proceeding with analyses, I examined the manipulation check to determine if the experimental manipulation was successful in invoking stereotype threat. T-tests identified a statistically significant difference between conditions ($t = 3.35$, $df = 126.41$, $p < .01$) such that those in the experimental condition reported that they were expected to do poorly on the task (see Appendix A8 for the full manipulation check measure). Thus, the manipulation appears successful in creating conditions of stereotype threat.

Mixed-effect ANOVAs were conducted as an omnibus test to examine whether there was significant variance in the variables of interest. Effort, Attributions, Feedback Usefulness, Goal Commitment, and Self-Efficacy all displayed group differences such that participants differed by experimental condition ($p <.05$). Attributions, Feedback Usefulness, Feedback Reactions, Goal

Commitment, and Self-Efficacy displayed a significant interaction term between condition and time, suggesting that variance between the experimental conditions varies across time ($p < .05$). Performance and Goal Level did not have significant condition nor condition*time interactions.

Trait measures of goal orientation (Learning, Prove, Avoid) and affect (Positive, Negative) were collected at the beginning of the experiment. State measures of these variables were collected intermittently (Blocks 1, 6, 11, 16, 21) to provide control measures of these variables. They were not collected every block to reduce participant burden. The mixed-effect ANOVA results from above were repeated as ANCOVAS including the goal orientation and affect measures as covariates. The only change in the reported results above was that the condition*time interaction for the Feedback Reactions scale was no longer significant ($p > .05$). Next, ANCOVA analyses were conducted only on the blocks for which state covariates were measured to determine if controlling for these covariates affected my conclusions. None of the statistical relationships changed meaningfully in this set of analyses.

Results from all 21 blocks were aggregated to examine whether there were overall differences in the focal constructs before conducting time-varying analyses. T-tests were conducted on each of the eight constructs, finding that Effort ($t = 2.58$, $df = 269.36$, $p < .05$), Attributions ($t = -4.95$, $df = 276.25$, $p < .01$), Feedback ($t = 4.01$, $df = 277.66$, $p < .01$) and Commitment ($t = 3.36$, $df = 260.26$, $p < .01$) displayed group differences. Overall, individuals in the experimental condition reported higher effort, performance attributions that were environmentally focused / stable / uncontrollable, more favorable attitudes about the usefulness of the feedback, and higher commitment to their self-set goals. Performance, Goal level, Feedback Reactions, and Self-efficacy were not significant at the aggregate level ($p < .05$).

Next, t-tests were performed at the Set level (Blocks 1-5 = Set 1, Blocks 6-10 = Set 2, Blocks 11-15 = Set 3, Blocks 16-21 = Set 4). Appendix B display results at both the 21-Block and 4-Set levels (see Appendix B2 footnote for coding/interpretation of these appendices). Analyses were organized into Sets to simplify analyses and reporting (see Analysis section); see Appendix C for a full listing of descriptive statistics and correlation tables across the 21 blocks. Beginning with Effort, Set 1 (t = 2.84, df = 262,59, p < .01); Set 2 (t = 2.37, df = 268.98, p < .05), and Set 4 (t = 2.50, df = 268.60, p <.05) were significant, with the experimental group displaying more effort than the controls. Set 3 was marginally significant (t = 1.91, df = 270.71, p = .06). This provides partial support to Hypotheses 1 (b) and 2 (b). As expected, those in the experimental group displayed more effort early on, and it does appear that effort may have converged between the two groups (although inferring this convergence relies on Set 3 narrowly falling below significance). However, the control group did not display more effort than the experimental group in later blocks as expected, so the overall cross-over pattern that was expected did not fully emerge.

Goal level never fell below p < .05, failing to provide support for Hypotheses 1 (a) and 1 (b). Similarly, the groups never differed in Performance, failing to provide support for Hypotheses 3 and 7.

Attributions differed across all four sets: Set 1 (t = -6.91, df = 290.05, p < .01); Set 2 (t = -4.46, df = 276.38, p < .01); Set 3 (t = -4.43, df = 275.51, p < .01); Set 4 (t = -3.40, df = 276.78, p < .01). Those in the experimental group were more likely to attribute performance to external, uncontrollable causes. This provides partial support for Hypothesis 5. While it was expected that attributions would only differ in later blocks, attributions differed immediately.

Feedback usefulness also differed across all four sets: Set 1 ($t = 3.35$, $df = 278.55$, $p < .01$); Set 2 ($t = 3.58$, $df = 279.78$, $p < .01$); Set 3 ($t = 3.70$, $df = 268.58$, $p < .01$); Set 4 ($t = 4.32$, $df = 270.72$, $p < .01$). Surprisingly, those in the experimental condition were more likely to rate the feedback they received as useful, rejecting Hypothesis 4 (b). Feedback reactions never achieved statistical significance, failing to provide support for Hypothesis 4 (a).

Goal commitment differed across all four sets: Set 1 ($t = 3.27$, $df = 261.68$, $p < .01$); Set 2 ($t = 3.02$, $df = 263.42$, $p < .01$); Set 3 ($t = 2.85$, $df = 264.79$, $p < .01$); Set 4 ($t = 3.43$, $df = 255.78$, $p < .01$). Across the experiment, those in the experimental condition reported higher commitment to their goals than those in the control condition. This provides support to Hypothesis 1 (c) and rejects Hypothesis 2 (c).

Finally, Self-efficacy varied across Set 1 ($t = 2.42$, $df = 237.28$, $p < .05$) and Set 2 ($t = 2.12$, $df = 242.75$, $p < .05$), with those in the experimental group reporting higher self-efficacy in early trials. There was no significant difference between the groups in Sets 3 and 4. This provides support for Hypothesis 6.

Latent growth curves were then generated in the lavaan R package (see Appendix D for sample R code). First, standard errors were extracted from the models and compared to the standard errors used in the t-test hypothesis testing to determine whether violation of statistical independence could contribute to erroneously finding significant results in that set of analyses. Analyses were replicated using these standard errors and no substantive differences were found. Next, a latent growth curve was generated for each of the eight focal variables. Interpretation of the simple change over time in these curves was not particularly important to these analyses; rather, differences in the slopes and intercepts based on Condition were examined by regressing slopes and intercepts on Condition.

Effort displayed a positive intercept difference (Estimate = -0.657, *SE* = .275, *z* = -2.39, *p* <.05) suggesting that those in the experimental group initially displayed more effort. No slope difference was found. Goal level displayed no slope nor intercept differences. Both the slope and intercept terms for Performance were marginal: intercept (Estimate = -1.527, *SE* = .837, *z* = 1.8, *p* = .07); slope (Estimate = .085, *SE* = .048; *z* = 1.77, *p* = .08). Attributions displayed both slope and intercept differences: intercept (Estimate = .916, *SE* = .147, *z* = 6.25, *p* < .01); slope (Estimate = -.02, *SE* = .007, *z* = -3.10, *p* < .01). Those in the experimental condition reported less favorable initial attributions than the control group, but those ratings tended to converge across time. Feedback usefulness displayed intercept differences (Estimate = -0.479, *SE* = .159, *z* = -3.00, *p* < .01) and a marginal slope difference (Estimate = -0.013, *SE* = .007, *z* = -1.935, *p* = .05) such that feedback usefulness was initially more favorable for the experimental group. Feedback reactions did not display slope nor intercept differences. Commitment displayed intercept differences (Estimate = -0.337, *SE* = 1.22, *z* = -2.77, *p* < .01) and a nonsignificant slope difference, indicating that those in the experimental group were initially more committed than those in the control group. Lastly, both the intercept (Estimate = -0.633, *SE* = .218, *z* = -2.902, *p* < .01) and slope (Estimate = .031, *SE* = .011, *z* = 2.816, *p* < .01) were significant, suggesting that self-efficacy was initially higher for those in the experimental group and tended to converge between the two groups across time. Overall, results from the latent growth curve analysis largely corresponded and clarified relationships described in the initial set of analyses.

DISCUSSION

Support for the proposed model was mixed. The proposed model suggested that those experiencing ST would initially report higher levels of motivation-related constructs on earlier trials of the experiment than the control group; on later trials, they would report lower levels compared to the control group. Feedback reaction/usefulness, self-efficacy, and attribution processes would drive this shift. Performance would likewise be higher initially and lower by the conclusion of the experiment, with suffering performance causing additional decrements in motivation.

Overall, those in the experimental group tended to report higher levels of motivation-related constructs (Effort, Goal Commitment). Contrary to hypotheses, those initial levels of motivation remained across the experimental session. Performance was attributed to environmental, uncontrollable causes by the experimental group across the experiment. Feedback was rated as more useful by the experimental group. Self-efficacy was higher for the experimental group across approximately the first half of the session. Performance, goal-level, and feedback reactions always exceeded $p = .05$ and did not display meaningful differences between the groups.

How might we interpret these results? First, we will begin with the variables that did not display experimental differences. Performance and goal level did not display meaningful group differences in the analyses conducted. One possible explanation is the complexity of the task. While participants did improve their performance as they learned the task, their stock estimates merely improved from within approximately $23 of the target in Set 1 to $19 in Set 4, and goals tended to be calibrated to performance. This presents a key challenge for stereotype threat research: stereotype threat is proposed to occur only when working memory is sufficiently taxed

(Schmader et al., 2008), but a highly complex task environment can inhibit the short-term learning and skill acquisition needed to produce performance differences. It is worth noting that stereotype threat studies frequently fail to find performance differences between conditions, even when sessions are hours long or spread across multiple sessions (Grand, 2017). Researchers should be careful to not conflate performance with motivation or other constructs and are cautioned to think carefully about the model and structure of performance (Campbell & Wiernik, 2015).

Feedback reactions also did not display differences, despite differences in feedback usefulness ratings. The feedback reactions scale used single-word emotion items to assess feedback reactions, while the feedback usefulness scales were short phrases (see Appendix A10). Weidman, Steckler, and Tracy (2017) noted that single-word items can be problematic for the assessment of distinct emotions due to the imprecision resulting from their brevity. It is plausible that the feedback reactions scale used here suffered from similar problems. Alternatively, the contrast in positive ratings of feedback usefulness relative to the nonsignificant differences in feedback reactions is interesting and deserves unpacking. The positive ratings of feedback usefulness by the experimental group were contrary to expectations. It is plausible that in light of their poorer performance attributions and waning self-efficacy, the feedback provided instrumental value that buffered downstream motivational and performance decrements. Individuals may have been able to compartmentalize their reactions and rely on feedback as a mechanism to overcome the suboptimal conditions of the experimental group.

Attributions were always rated by the experimental group as due to environmental, uncontrollable causes. In hindsight, it may be consistent with the chronic threat model that these attributions would always be in this direction following the experience of stereotype threat.

Individuals may not need significant task exposure to confirm these attributions, particularly when a ST manipulation is blatant about their expected level of performance.

Self-efficacy displayed encouraging results for the model. Self-efficacy was higher initially for controls, and displayed no group differences by the conclusion of the experiment. A full cross-over effect where controls rated their self-efficacy as higher by the conclusion was not observed, but this seems reasonable given the time-limited nature of the study. Of further interest is that self-efficacy dropped despite performance remaining constant between the two groups. It is extremely promising for the model that ST effects could begin to emerge absent veritable differences in performance, which could eventually result in downstream reduction in motivation and eventual performance decrements.

It appears that the manipulation successfully induced individuals in the experimental group to work harder to achieve their goals. Motivation, however, did not drop below the control group in the latter stages of the experiment. As suggested above, feedback reaction/usefulness, attribution, self-efficacy, and performance decrement processes were all expected to reduce motivation across time. As expected, individuals in the experimental group were more likely to attribute performance to external, uncontrollable, permanent circumstances rather than make internal, controllable, and malleable attributions. This was present immediately after the manipulation at the beginning of the experiment. While individuals were able to maintain their motivation, there appears to be some recognition and reduction in self-efficacy following these attributions. These factors may not have been sufficient to cause motivational or performance decrements in a single lab session but provides some tentative evidence for the conceptual model presented. Relatedly, repeated exposure to ST across time may cause feedback reactions and usefulness to orient towards frustration; if attributions are external and feedback is not viewed as

useful or positively-valent, the combination of these factors may be the inus conditions necessary for reduced motivation and task withdrawal.

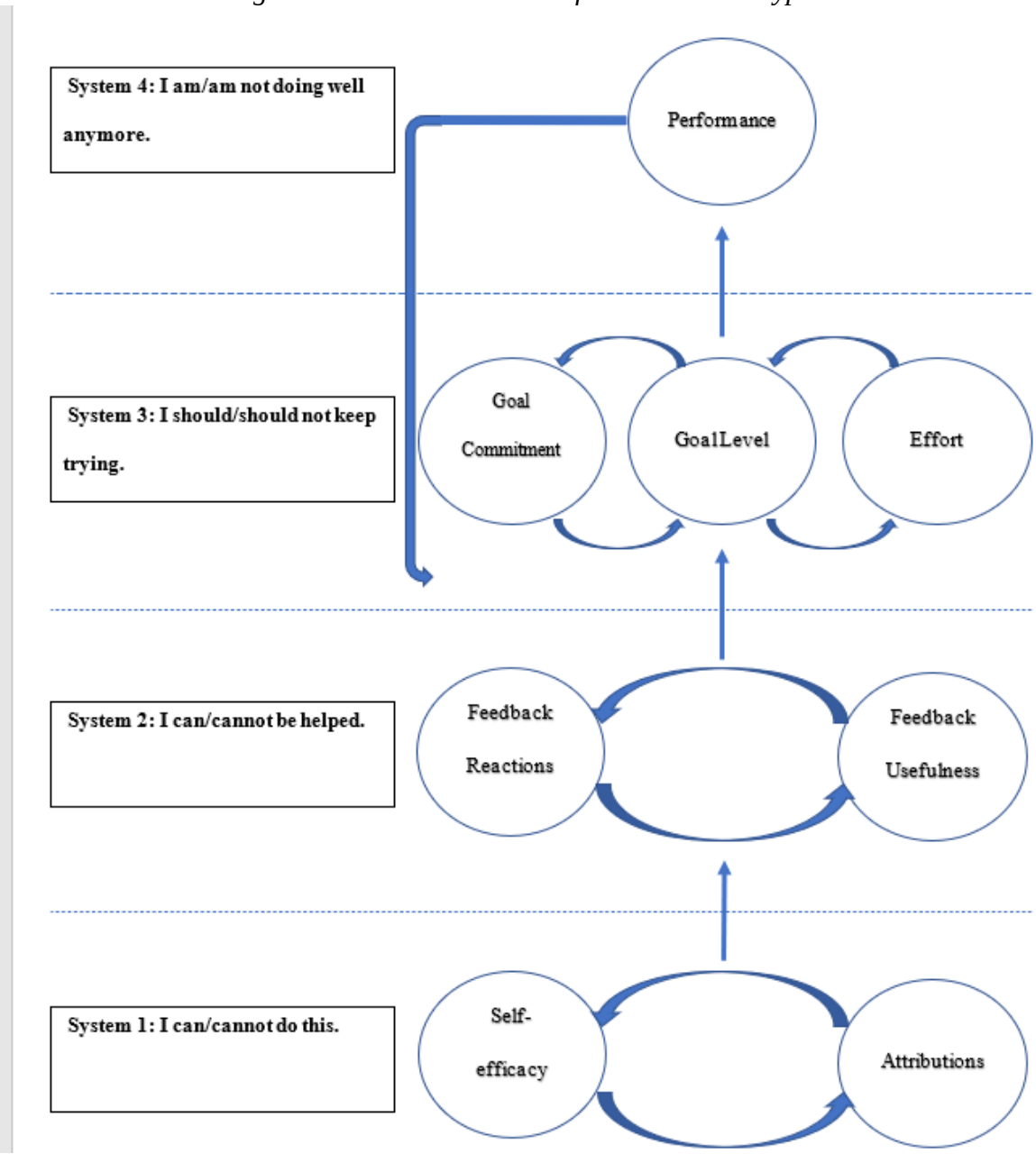
Overall, it appears that stereotype threat affects more proximal constructs (self-efficacy, attributions) and may require longer periods of time for those effects to propagate towards constructs like goals or performance. Indeed, the chronic threat model of stereotype threat presented here is meant to simulate what happens across repeated exposure to threat within weeks, months, and years. It is encouraging that several aspects of the model were supported and that many of the focal constructs examined displayed meaningful differences across conditions within a relatively short lab session. Taken together, the initial model of stereotype threat may be appropriate and simply require further observation across time to be fully realized.

Otherwise, the model could be respecified. Figure 1 implies that the constructs specified in the model cyclically and steadily influence each other across time. Individuals set goals, commit to and effortfully pursue their goals, achieve some level of performance, receive feedback, interpret feedback, make attributions about this feedback, and experience changes in self-efficacy. Figure 1 suggests that each of these constructs changes on a similar sequential timescale. This performance cycle feeds into the next cycle and causes steady downstream deficits in the focal constructs. However, the observed data tentatively suggest that the effects of stereotype threat may instead be more immediate for some aspects of the model, while other aspects may require repeated exposure to stereotype threat to manifest.

See Figure 2 for an alternative model of chronic stereotype threat. In this model, I have separated the focal constructs of stereotype threat into four systems. The lowest level of this model, System 1, displays proximal constructs that are affected relatively immediately by stereotype threat. System 1 is initialized as one's perception that "I can do this / I have control."

Prior to any exposure to threat, individuals tend to believe that they have control and are capable of acting upon their world (Bandura, 1986). However, we observed in our data that attributions shifted immediately following stereotype threat, and self-efficacy required just half of an experimental session to shift. This system appears to be most malleable following threat, eventually shifting to a perception of "I cannot do this / I do not have control."

Figure 2: Alternative Model of Chronic Stereotype Threat



Even if an individual does not perceive that they have control nor that they can perform, they may still recognize that they could be helped or that others could change the situation. Indeed, one of the dimensions of performance attributions indicates whether a situation is controlled by oneself or governed by others. System 2 begins with the perception that "I can be helped." At this stage, it is expected that continuous exposure to threatening conditions will cause feedback reactions to shift to a negative valence and feedback will be perceived as less useful. This shift was not observed in the current study, suggesting that this system is more distal than the variables specified in System 1. Though feedback in this experiment was expected to be interpreted negatively and to make the performance cycle in Figure 1 especially salient, feedback may have helped buffer stereotype threat effects in the short-term of the experimental session. The alternative model of chronic stereotype threat suggests that given sufficient exposure to threat, this system would eventually shift to the state of "I am helpless."

System 3 begins with the perception that "I should keep trying." If an individual believes that they have control (System 1) or that they can be helped (System 2; even if they do not have control), it follows that a task is worth engaging effort toward where it provides some instrumental value (Vroom, 1964; Kukla, 1972). Disrupting both of these systems creates a scenario where the individual does not believe that they can achieve their goals even if support systems are in place. When an individual lacks sufficient belief that they can achieve their goals, both self-efficacy theory (Bandura, 1986) and cybernetic theories of self-regulation (Vancouver et al., 2001) agree that the individual will withdraw effort ("I should not keep trying"). Motivation-related constructs (e.g., goal level, goal commitment, effort) are located here and are expected to be stable due to their elevated position in the hierarchy. Indeed, System 1 and

System 2 will need to be especially affected to overcome the initial "I'll prove you wrong" motivational boost predicted by stereotype threat research.

System 4 describes the ultimate performance outcomes of withdrawing effort. This system initializes as "I am doing well." This is conceptually distinct from subjective self-efficacy beliefs; recall that performance differences were not observed despite changes in self-efficacy across the experimental session. Even when an individual believes that they (1) do not have control, (2) cannot be helped, and (3) should keep trying, effects on performance may not be immediate. Performance is unimpeded throughout the more proximal systems of the model. It is only when an individual displays withdrawal behavior from the domain in System 3 that they will begin to suffer performance decrements and shift towards the final state in the model of "I am not doing well." Attitudes and perceptions do not have performance consequences until they manifest into behaviors.

Finally, diminished performance outcomes are expected to feed back to lower levels of the system. This feedback articulates how stereotype threat may create a self-fulfilling prophecy through which performance worsens and motivation is further withdrawn. The alternative model tentatively restricts the feedback as only affecting System 3 rather than looping all the way back to System 1. The model as described requires crystallization within systems before perpetuating to the next system. Individuals must believe that they cannot perform and do not have control before they believe that they also cannot be helped by others; they believe that they should not try only after both internal and external avenues to achieving their goals are exhausted. When performance is ultimately reduced in System 4, it may not be meaningful to suggest that attributions are "more" externally focused nor that feedback is "less" useful when performance outcomes are temporally far-removed from these earlier systems. Instead, performance deficits

signal that individuals are indeed not achieving their goals and that effort should be further withdrawn from the domain.

Future research will be required to further test both the original and alternative versions of the chronic threat model. Repeated measures will be required with careful mapping of when threat occurs. An important limitation of the current research was that the model is not yet refined enough to specify "when" threat becomes chronic. Organizing the analyses into blocks was intuitive, and we may expect that repeated exposures to threat have a probabilistic association with eventually experiencing chronic threat, but is threat steadily cumulative (Figure 1) or is there a breakpoint, a discontinuity, where threat becomes chronic (Figure 2)? Indeed, the current operationalization and theoretical development treats chronic stereotype threat as a phenomenon that emerges from quantitative exposure to threat. A reasonable alternative is that chronic stereotype threat is a qualitatively different phenomenon from acute stereotype threat, emerging from factors other than simply the amount or frequency of threat experienced. For example, chronic threat could emerge from experiencing different threatening cues across time (e.g., underrepresentation in the workplace/classroom, identity-salient cues in the physical workplace, diagnostic/high-stakes contexts). More work is needed to fully understand the nature of chronic threat.

Relatedly, Shapiro, Williams, and Hambarchyan (2013) discovered that subtleties in the nature of a stereotype threat manipulation affected the efficacy of interventions designed to mitigate threat. Specifically, they distinguished between group-as-target and self-as-target stereotype threat. Group-as-target threat occurs when an individual's ingroup is the target of the stereotype, and failing to perform would confirm stereotypes about their ingroup. Self-as-target threat occurs when an individual is concerned about their own performance, and failing to

perform confirms that their own performance suffers due to their group membership. Their specific manipulations read as follows (group-as-target and self-as-target, respectively):

"In today's session, we want to get a measure of intellectual ability for Black and White students by having you take a standardized intelligence test . . ..Your performance on this test will be used to help us establish intellectual performance norms for Black and White students. After the test, we will provide you with feedback about Blacks' performance relative to Whites' performance and ask you some questions about test taking." (p.280)

"In today's session, we want to get a measure of your intellectual ability by having you take a standardized intelligence test . . ..Your performance on this test will be used to help us establish your personal intellectual ability. After the test, we will provide you with feedback about your performance relative to other students and ask you some questions about test taking." (p.280)

My manipulation (unedited) is displayed below:

"This task has been designed to examine possible explanations for why women tend to perform more poorly than men on tasks involving mathematical and analytic aptitude. **One reason for this finding may be that women have more difficulty than men distinguishing relevant information needed to solve a problem.** This task has been specifically designed to examine differences in these skills. Individuals who do not have the ability to distinguish important problem-relevant information will struggle to perform well on this task."

This manipulation may have aspects of both self-as-target and group-as-target threat. In one sense, it is implied that individuals are expected to perform more poorly based on their group membership (self-as-target). On the other hand, it is blatant that group membership is the basis for this pessimistic performance outlook, and individuals may feel that they need to work hard to

advocate for their group (group-as-target). The nature of the manipulation may thus be important because the motivational potential of stereotype threat may be motivating when they need to advocate for their group (group-as-target) and demotivating when the individual is unlikely to perform well due to their group status (self-as-target). Specific to the current study, one could hypothesize that the group-as-target threat is indicative of acute Stereotype Threat and the self-as-target threat indicates chronic threat.

Follow-up research could try to tease this apart, as several sequences of events are possible. First, individuals exposed to the same threat conditions/manipulations could come to reframe the threat as self-targeted or group-targeted as they are repeatedly exposed to it, especially if it is somewhat open-ended as in the current study. Perhaps chronic threat manifests when that shift occurs. Alternatively, one may need to experience qualitatively different threats, referred to by Shapiro et al. (2013) in their Multi-Threat Framework, for threat to become chronic. In the current conceptualization, that would necessitate experiencing group-as-target acute threats followed by self-as-target threats that qualitatively shift the threat to chronic. In summary, the nature of when and how Stereotype Threat shifts to a chronic nature remains unclear, but several avenues exist for further research.

Limitations

Although self-efficacy was included in the chronic threat model, several researchers suggest that ST causes individuals to disengage their self-esteem from the domain or task environment (Crocker et al., 1998; Steele, 1997). However, there are unresolved issues of temporal and conceptual clarity. The process of "disengaging self-esteem" is somewhat nebulous and may conflate self-efficacy and motivation, which I attempted to separate in the current research both in measurement and in timescale (Figure 2). Stereotype threat researchers may, however, deem these processes inseparable and co-occurring.

Relatedly, there are a flurry of terms used to describe threat-relevant attitudes and behaviors; a small sampling includes "withdrawal," "(dis)identification," "disengagement," "interest," and "self-esteem," as well as varied permutations of these terms. Although distinctions between some terms have been articulated, the literature may be susceptible to "jingle/jangle" fallacies (Kelley, 1927), where the same label may be used for discrete constructs ("jingle") and different labels are used to identify a single construct ("jangle"). The ST literature may be especially susceptible to these fallacies as researchers from various disciplines continue to enter the dialogue, inserting their preferred terminologies. Additional conceptual and empirical work is required to more precisely delineate the stereotype threat space; a comprehensive framework and a common language is needed.

My method of introducing chronic threat may differ from the chronic threat experienced in operational settings. I used a stereotype threat manipulation to induce an initial state of acute threat, followed by "gender checks" meant to revive the threat prior to each block. However, ST in an operational setting will likely differ from these blatant initiations of threat, and also may differ qualitatively across time. Individuals often experience threat in multiplicative, varied

streams, such as through group underrepresentation, organizational messaging, and even the décor of the physical space (Walton et al., 2015). It is possible that these subtler stereotype threats may be insufficient to provoke chronic, recursive consequences. On the other hand, myriad constant environmental reminders of threat may inhabit the individual's psychological space more readily than the current manipulation. Further research will need to explore how chronic threat may manifest and persist in organizational settings.

Finally, presentation of the alternative chronic threat model was motivated by the failure of the study to instantiate motivational and performance decrements in the experimental group. I suggested that stereotype threat is conceptualized as a phenomenon that occurs over vast periods of time (Steele, 1995), resulting in chronic disengagement from the domain. The single-session lab study collected here, despite its repeated-measures design, appears to have been insufficient to capture the full chronic threat process. This study serves as proof of concept for the initial conditions of chronic stereotype threat activation, but will require truly longitudinal work to explicate.

Future Directions

Introducing goal regulation as an explicit framework for motivation-based stereotype threat research is an important development for the literature both due to its empirical connection to real-world self-regulation (Ackerman et al., 2017) and for its strength over trace- and outcome-based measures of motivation (recall Table 1). Future ST researchers should continue to clarify the interactive cognitive, motivational, and affective pathways through which ST may impair performance. For example, resource depletion is an active area of research in the self-regulation literature (Kanfer et al., 2017). Muraven and Baumeister (2000) provided an early

model of this theory, proposing that self-regulation results in expenditure of inner resources. Ego-depletion effects have suffered from failed replications (e.g., Hagger et al., 2016), but appear to offer enough appeal that they have not been dismissed from self-regulation research (see Kanfer et al., 2017). Resource depletion perspectives, or alternative resource-based conceptualizations, could offer specific promise for ST research to the extent that they delineate how cognitive, motivational, and affective systems are taxed across time, perhaps separately, due to experiencing threat. For example, effortful activities such as monitoring the task environment for possible problems (Lin & Johnson, 2015) and faking and suppressing emotional displays (Trougakos, Beal, Cheng, Hideg, & Zweig, 2015) appear to result in depletion of self-regulatory resources (Kanfer et al., 2017). Future research could explore whether experiencing stereotype threat is especially resource depleting. The finding that monitoring for problems is resource depleting (Lin & Johnson, 2015) hints that ST may be depleting, as those experiencing stereotype threat engage in more active monitoring (Schmader et al., 2008).

Researchers could explore threatened individuals' reactions to different forms of feedback, such as those introduced in the typology by Kozlowski, Toney, Mullins, Weissbein, Brown, & Bell (2001). Kozlowski et al. (2001) describe three interpretation properties of feedback: evaluation, attribution, and guidance. These properties describe how individuals interpret *what* happened, w*hy* it happened, and *how* the individual needs to adjust their future task strategies (p. 68-69). It's plausible that feedback effects within stereotype threat can be better understood by considering how individuals interact with different components of feedback. Relatedly, Ilies and Judge (2005) manipulated feedback framing so participants received either feedback relative to their personal goals and performance (nominal feedback) or feedback relative to other participants (relative feedback). Their theoretical rationale for

manipulating this interpretive frame (the evaluation component in the Kozlowski et al. taxonomy) was that this feedback framing could influence performance attributions: comparative feedback is affected by both one's personal performance as well as the external, uncontrollable performance of others, while nominal feedback does not contain this comparator. Within a stereotype threat study, we might expect that relative feedback would make stereotype threat especially salient, particularly if the referent group contained individuals that are not stereotypically expected to perform poorly.

Stereotype threat could also be explored within a multiple-goal regulation context. There has been significant research attention recently examining how individuals concurrently regulate attention across multiple goals. A primary motivation for this has been the realization that single-goal paradigms do not necessarily clarify how individuals self-regulate across more than one domain (Unsworth, Yeo, & Beck, 2014). One of the first articles to show this was published by Kernan and Lord (1990), who showed that valence and expectancies were important for goal prioritization and resource allocation in a multiple-goal task, but were unimportant for single-goal pursuit. Stereotype threat could affect multiple-goal regulation in a few ways. For example, the increased self-monitoring of threatened individuals may distract them from adequately tracking goal progress and allocating attention appropriately. Additionally, it would be of interest to examine how individuals allocate attention across several goals when they are experiencing stereotype threat and some of the available goals are stereotyped domains. If individuals were less likely to pursue goals from stereotyped domains, it would provide a compelling laboratory model for the domain abandonment commonly found in longitudinal studies.

The consequences of stereotype threat for adaptive learning outcomes have been recently explored by Grand (2017), primarily from a cognitive perspective. Further research could

examine the motivational consequences of stereotype threat for adaptive performance. As mentioned, Jamieson and Harkins (2007, 2009, 2011) found that threatened individuals tend to adopt prepotent response strategies. It is plausible that these strategies may facilitate routine performance but could be deleterious for performance in dynamic situations. Adaptive performance has received increased attention as organizations have trended towards requiring individuals with malleable capabilities rather than routine expertise (Baard, Rench, & Kozlowski, 2014; Bell & Kozlowski, 2008). Thus, inflexibility caused by ST could be especially problematic in operational settings. Carr and Steele (2009) have provided some initial evidence for this hypothesis, finding that threatened individuals are more likely to engage in "inflexible perseverance," failing to adjust task strategies when they are no longer effective.

Implications

The chronic threat model should be informative for stereotype threat researchers who wish to anticipate whether stereotype threat will cause stereotype reactance or withdrawal behavior. Researchers should explicitly assess the temporal and structural elements of their study. They should consider whether participants are constantly exposed to the stereotype physically (e.g., group representation), psychologically (e.g., the task environment makes the stereotype readily accessible), or otherwise.

Even if chronic threat is conceptualized as requiring repetitive exposure to negatively impact motivation and performance, more proximal constructs (e.g., attributions) were affected immediately. It remains unclear how stereotype threat may cause differences in practical settings (Shewach et al., 2018), but the relative immediacy through which many variables seem to be altered by ST should caution practitioners against allowing ST-invoking cues to be present in the

workplace. Both models of threat presented here do not necessarily require the cues to be persistently salient for chronic threat to occur. Stereotype threat initializes an interrelated constellation of thoughts and actions that could perpetuate absent repeated threats.

Likewise, the downstream effects of ST could have implications for criteria other than job or task performance. Figure 2 frames performance as the most distal outcome of stereotype threat, while factors such as attributions and self-efficacy may require little to undermine. Variables such as attitudes (e.g., job satisfaction) and behaviors (e.g., turnover, CWBs) may suffer before differences in performance are observed and these likewise have organizationally-relevant consequences. Steele's original (1995) proposition is more akin to a model of turnover than of performance and is worth reflection. Criteria should be calibrated appropriately.

CONCLUSION

A chronic threat model of Stereotype Threat (ST) was presented to bridge discrepant findings that emerged in the literature. ST was proposed to provide initial motivational energy that is eventually undermined by attribution, feedback, and self-efficacy mechanisms. While several of the focal variables displayed expected relationships with the experimental condition, performance and motivational mechanisms largely remained unchanged. The current study provides some tentative evidence for the elementary conditions that may result in chronic stereotype threat, but requires study across longer time periods to fully capture the transition from acute experiences of threat to chronic threat.

APPENDICES

APPENDIX A:
Survey Measures

Appendix A1:
Demographics (Questions administered online using dropdown menus).
Gender-inclusive measures from Bauer et al. (2017)

Demographics Questionnaire

Please provide as much of the following information as is applicable. It is important to understand that these scores will be kept confidential and used only for research purposes. If you do not remember your exam scores, please put a zero in that space.

What sex were you assigned at birth, meaning on your original birth certificate?

1. Male
2. Female

Which best describes your current gender identity?

1. Male
2. Female
3. Indigenous or other cultural gender minority (e.g. two-spirit)
4. Something else (e.g. gender fluid, non-binary)

Age: ____

What is your ACT or SAT score? (If you did not take these or don't remember, write "0")
Math _____ / Verbal _____

Did you complete the MSU Math Placement Exam when you entered MSU? Yes/No

If yes, which mathematics course did you place into upon entering MSU? (MTH 1825; MTH 101; MTH 102; MTH 103; MTH 110; MTH 116; MTH 124; MTH 201; STT 200; STT 201; MTH 132; Don't remember).

Year in College (using SBEI drop down menu):

Major (using SHEI drop down menu):

What is your ethnicity?
[ ] Hispanic or Latino
[ ] Not Hispanic or Latino

Race (Check all that apply):
[ ] White

[ ] Black or African American

[ ] American Indian or Alaska Native

[ ] Asian

[ ] Native Hawaiian or Other Pacific Islander

Is English your primary language (Yes/No)

Appendix A2:
Trait Goal Orientation.
From VandeWalle (1997)

For each of the following statements, please indicate how true it is for you on the scale provided below.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Goal Orientation Prove:

1. I prefer to do things that require a high level of ability and talent.
2. I'm concerned with showing that I can perform better than my peers.
3. I try to figure out what it takes to prove my ability to others.
4. I enjoy it when others are aware of how well I am doing.
5. I prefer to participate in things where I can prove my ability to others.

Goal Orientation Avoidance:

1. I would avoid taking on a new task if there was a chance that I would appear incompetent to others.
2. Avoiding a show of low ability is more important to me than learning a new skill.
3. I'm concerned about taking on a task if my performance would reveal that I had low ability.
4. I prefer to avoid situations where I might perform poorly.

Goal Orientation Learning:

1. I am willing to take on challenges that I can learn a lot from.
2. I often look for opportunities to develop new skills and knowledge.
3. I enjoy challenging and difficult activities where I'll learn new skills.
4. For me, development of my abilities is important enough to take risks.

Appendix A3:
State Goal Orientation.
From Horvath et al. (2001)

For each of the following statements, please indicate how true it is for you on the scale provided below.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree

Goal Orientation Prove:

1. It is important to me to perform better than others on this task.
2. It is important to me to impress others by doing a good job on this task.
3. I want the experimenter and other students to recognize that I am one of the best on this task.
4. I want to show myself how good I am on this task.
5. On this task, my goal is to perform well.

Goal Orientation Avoid:

1. On this task, I would like to hide from others that are better than me.
2. On this task, I would like to avoid situations where I might demonstrate poor performance.
3. On this task, I would like to avoid discovering that others are better than me.
4. I am reluctant to ask questions about this task because others may think I'm incompetent.
5. On this task, my goal is to avoid performing poorly.

Goal Orientation Learning:

1. I prefer to work on aspects of this task that force me to learn new things.
2. I am willing to work on challenging aspects of this task that I can learn a lot from.
3. The opportunity to learn new things about this task is important to me.
4. The opportunity to work on challenging aspects of this task is important to me.
5. On this task, my goal is to learn the task as well as I can.

Appendix A4:
Goal Commitment.
From Klein et al. (2014)

For each of the following statements, please indicate how true it is for you on the scale provided below.

1	2	3	4	5
Not at All	Slightly	Moderately	Quite a bit	Extremely

1. How committed are you to your goal?
2. To what extent do you care about your goal?
3. How dedicated are you to your goal?
4. To what extent have you chosen to be committed to your goal?

Appendix A5:
Positive and Negative Affect Schedule Short Form.
From Thompson (2007)

Question: Thinking about yourself and how you normally feel, to what extent do you generally feel:

Items in order:
Upset
Hostile
Alert
Ashamed
Inspired
Nervous
Determined
Attentive
Afraid
Active
Interval measure: *never* 1 2 3 4 5 *always*

Appendix A6:
Shortened State PANAS.
From Koy and Yeo (2008) and Yeo et al. (2014)

0	1	2	3	4	5	6	7	8	9	10
Not at All										Extremely

Please indicate the extent to which you experienced each of the following affective states during the previous block:

Items in order:

Enthusiastic

Upset

Interested

Distressed

Determined

Nervous

Appendix A7:
Mathematic Subscale of the Domain Identification Measure.
From Smith and White (2001)

Using the following scale, please indicate the number that best describes how much you agree with each of the statements below.

1	2	3	4	5
Strongly Disagree	Moderately Disagree	Neither Disagree or Agree	Moderately Agree	Strongly Agree

1. _____ Mathematics is one of my best subjects.
2. _____ I have always done well in Math.
3. _____ I get good grades in Math.
4. _____ I do badly in tests of Mathematics.

Please indicate the number that best describes you for each of the statements below using the following scale:

1	2	3	4	5
Not at All		Somewhat		Very much

5. _____ How much do you enjoy math-related subjects?
6. _____ How likely would you be to take a job in a math-related field?
7. _____ How much is Math to the sense of who you are?
8. _____ How important is it to you to be good at Math?
9. _____ Compared to other students, how good are you at math?

Appendix A8:
Manipulation Check.
Adapted from Bergeron et al. (2006); Grand (2017); Steele and Aronson (1995)

For each of the following statements, please indicate how true it is for you on the scale provided below.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

1. The researchers expected me to do poorly on this task because of my gender.
2. I doubt that others would think I have less ability on this task because of my gender.
3. Some people feel I have less ability on this task because of my gender.
4. The task may have been easier for people of my gender.
5. My gender does not affect people's perceptions of my ability.

Appendix A9:
Self-Set Goal Level

Goal Level

Please indicate your desired level of performance during the next block of trials:

My goal is to make predictions within $_____ of the actual stock values during the next block of trials.

(As a reminder, stock values range from $10 to $150. We suggest setting a goal of making predictions within $20 of the actual stock values).

Appendix A10:
Feedback Reactions and Usefulness.
From Brett and Atwater (2001)

Please indicate the extent to which you feel this way now based on the feedback you received (0 = *not at all*, 4 = *very much*):

1. Inspired 0 1 2 3 4
2. Angry 0 1 2 3 4
3. Encouraged 0 1 2 3 4
4. Judged 0 1 2 3 4
5. Confused 0 1 2 3 4
6. Informed 0 1 2 3 4
7. Aware 0 1 2 3 4
8. Examined 0 1 2 3 4
9. Pleased 0 1 2 3 4
10. Criticized 0 1 2 3 4
11. Motivated 0 1 2 3 4
12. Discouraged 0 1 2 3 4
13. Enlightened 0 1 2 3 4

This feedback was useful to me: 0 1 2 3 4

This feedback is useful for helping me diagnose my performance abilities: 0 1 2 3 4

Appendix A11:
Performance Attributions.
From McAuley et al. (1992)

Instructions: Think about the reason or reasons for your performance during the last block of trials. The items below concern your impressions or opinions of this cause of causes of your performance. Circle one number for each of the following questions.

Is the cause(s) something:

1.	That reflects an aspect of yourself	9 8 7 6 5 4 3 2 1	reflects an aspect of the situation
2.	Manageable by you	9 8 7 6 5 4 3 2 1	not manageable by you
3.	Permanent	9 8 7 6 5 4 3 2 1	temporary
4.	You can regulate	9 8 7 6 5 4 3 2 1	you cannot regulate
5.	Over which others have control	9 8 7 6 5 4 3 2 1	over which others have no control
6.	Inside of you	9 8 7 6 5 4 3 2 1	outside of you
7.	Stable over time	9 8 7 6 5 4 3 2 1	variable over time
8.	Under the power of other people	9 8 7 6 5 4 3 2 1	not under the power of other people
9.	Something about you	9 8 7 6 5 4 3 2 1	something about others
10.	Over which you have power	9 8 7 6 5 4 3 2 1	over which you have no power
11.	Unchangeable	9 8 7 6 5 4 3 2 1	changeable
12.	Other people can regulate	9 8 7 6 5 4 3 2 1	other people cannot regulate

Appendix A12:
Self-Efficacy

Please circle the percent likelihood that you will be able to estimate stock prices at the levels specified below during the next block of trials:

1. Within $50 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
2. Within $45 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
3. Within $40 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
4. Within $35 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
5. Within $30 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
6. Within $25 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
7. Within $20 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
8. Within $15 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
9. Within $10 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
10. Within $5 of the actual stock prices: 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Appendix A13:
Postexperimental Questionnaire.
From Earley et al. (1989) (except open-responses)

1. After you familiarized yourself with how to work on the task (say, after the first five predictions), how often did you try new strategies or methods for making stock predictions? (Check one)
    a. ___ I tried one or two methods and then stayed with a single one.
    b. ___ I tried several new ways to see which one was best.
    c. ___ I kept trying new ways of making predictions throughout the experiment and finally settled on one.
    d. ___ I tried new ways to make predictions throughout the experiment and would switch back and forth between several effective methods.
    e. ___ Other (please specify) _____

2. How useful were each of the three cues in helping you make predictions (1 = *not at all useful,* 5 = *extremely useful)*
    a. Marketing division performance: 1 2 3 4 5
    b. Research and development division performance: 1 2 3 4 5
    c. Production division performance: 1 2 3 4 5

3. Please rate each cue in terms of how much it helped you make predictions (1 = *poor,* 5 = *excellent)*
    a. Marketing division performance: 1 2 3 4 5
    b. Research and development division performance: 1 2 3 4 5
    c. Production division performance: 1 2 3 4 5

4. Allocate 100 points among the three cues to indicate the relative weight allotted to each
    a. _____ Marketing division performance
    b. _____ Research and development division performance
    c. _____ Production division performance
        i. (*The number of points given to the three cues must add up to 100 to advance)*

5. What kind of strategies did you use to regulate your emotions and adjust your prediction methods during this experiment?
6. How did you generally react to receiving feedback during the experiment?
7. How do you think the feedback could have been structured to help you regulate your emotions and adjust your prediction methods?

APPENDIX B:
Plots and Figures

Figure B1:
Example of Expected Motivation Plot

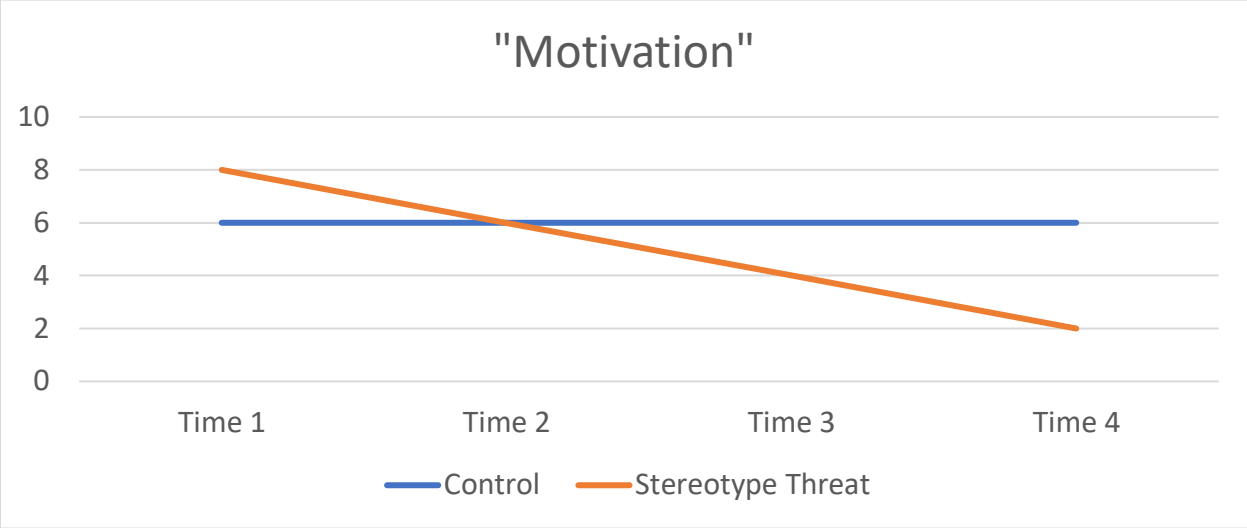
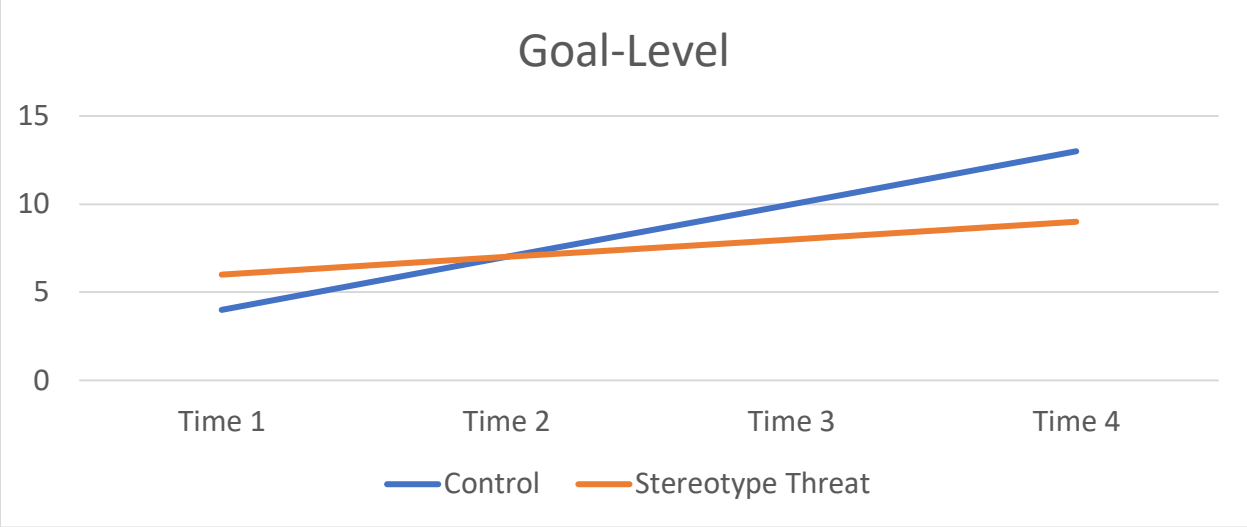


Figure B2:
Example of Expected Goal-Level Plot



[3]

[3] To make interpretation easier, the "Goal-level" graph is scored such that a goal of "estimate within $5" is a higher goal than "estimate within $10." The X and Y Values are imprecise and for illustrative purposes only. Goal-level increases as individuals improve on the task, but increases slower for threatened individuals. The "Motivation" graph displays a heuristic of motivation since goal-level should also be influenced by learning effects. Our effort measure may approximate this graph.

Figure B3:
Effort Plot by 21 Blocks[4]

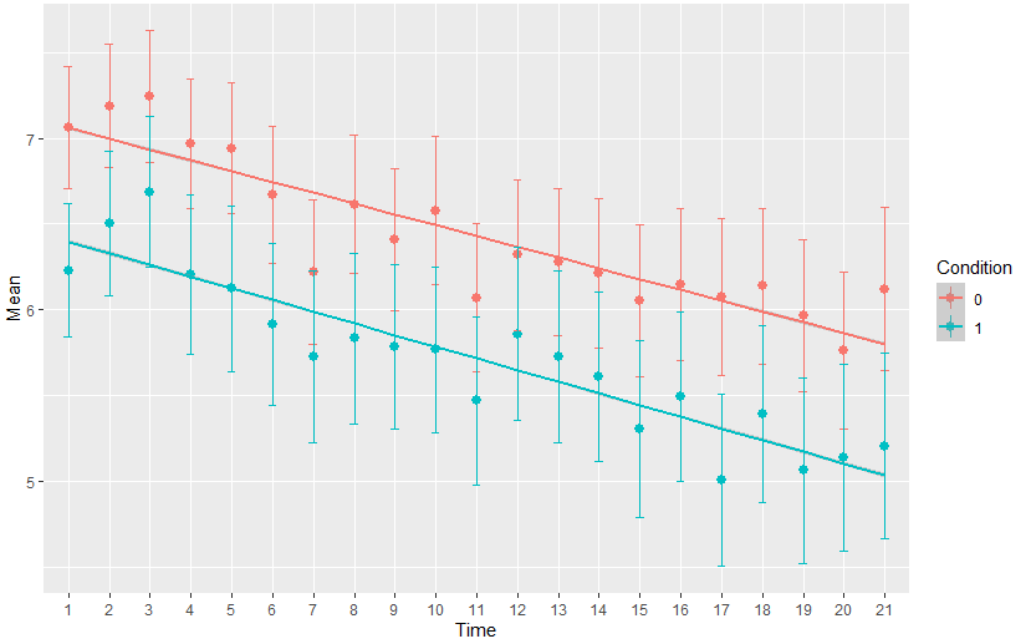
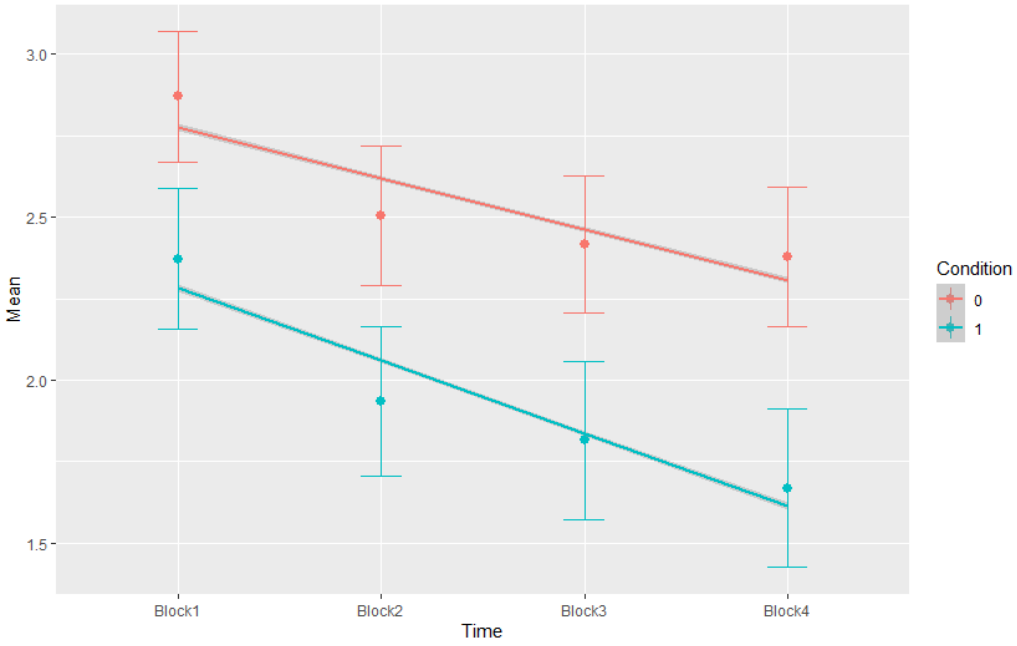


Figure B4:
Effort Plot by 4 Sets[5]



[4] These plots are all coded as 0 = Experimental, 1 = Control. The top plots are organized by the 21 original blocks, and the bottom plots collapse across Blocks 1-5, 6-10, 11-15, and 16-21.
[5] These plots are all coded as 0 = Experimental, 1 = Control. The top plots are organized by the 21 original blocks, and the bottom plots collapse across Blocks 1-5, 6-10, 11-15, and 16-21.

Figure B5:
Goal Plot by 21 Blocks[6]

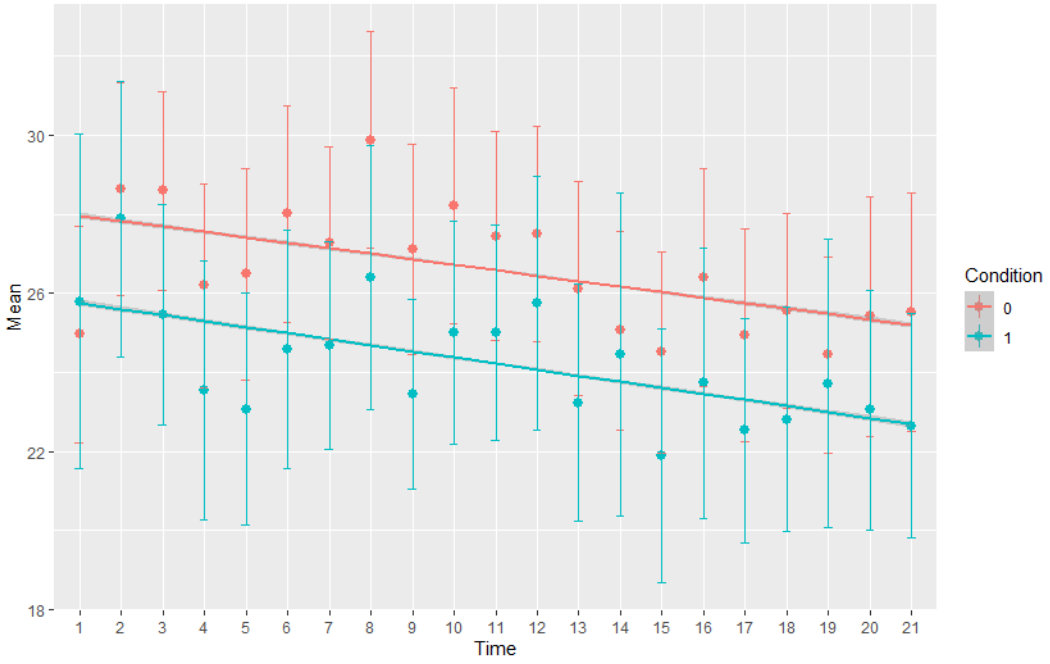
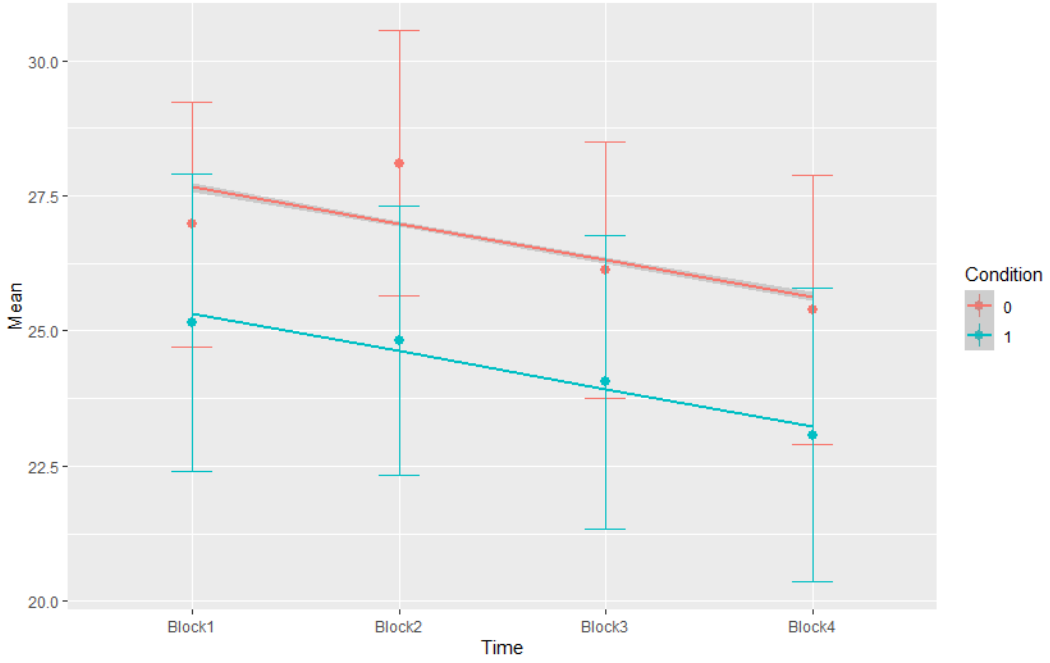


Figure B6:
Goal Plot by 4 Sets[7]



[6] Note that lower goal level = more ambitious guesses to the actual stock price and stronger goals
[7] Note that lower goal level = more ambitious guesses to the actual stock price and stronger goals

Figure B7:
Performance Plot by 21 Blocks[8]

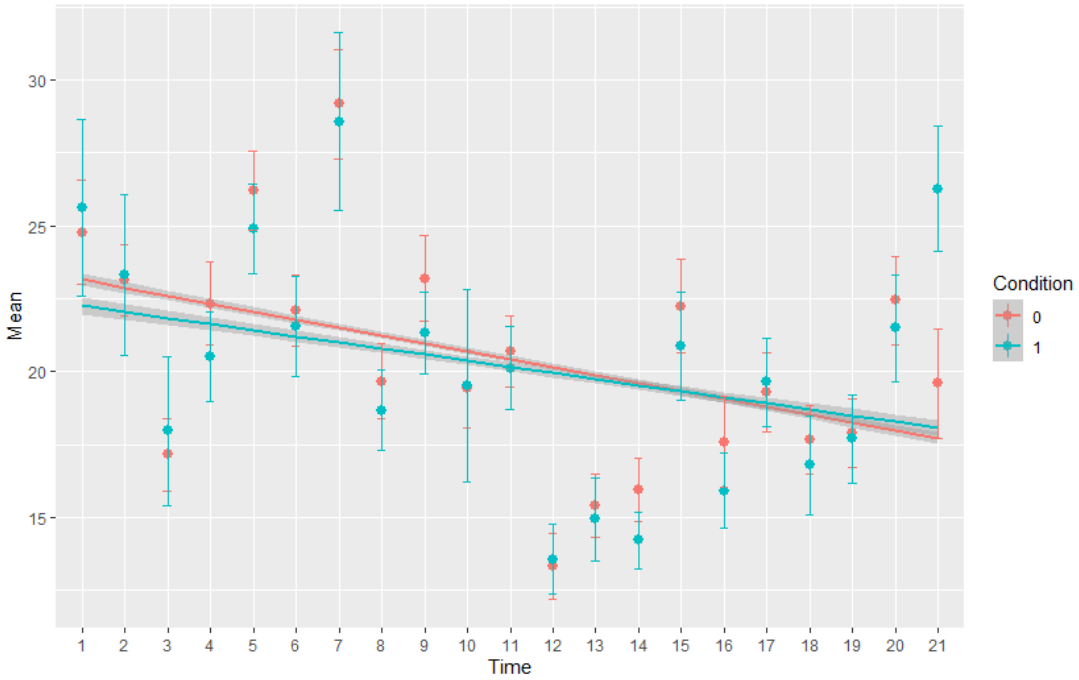
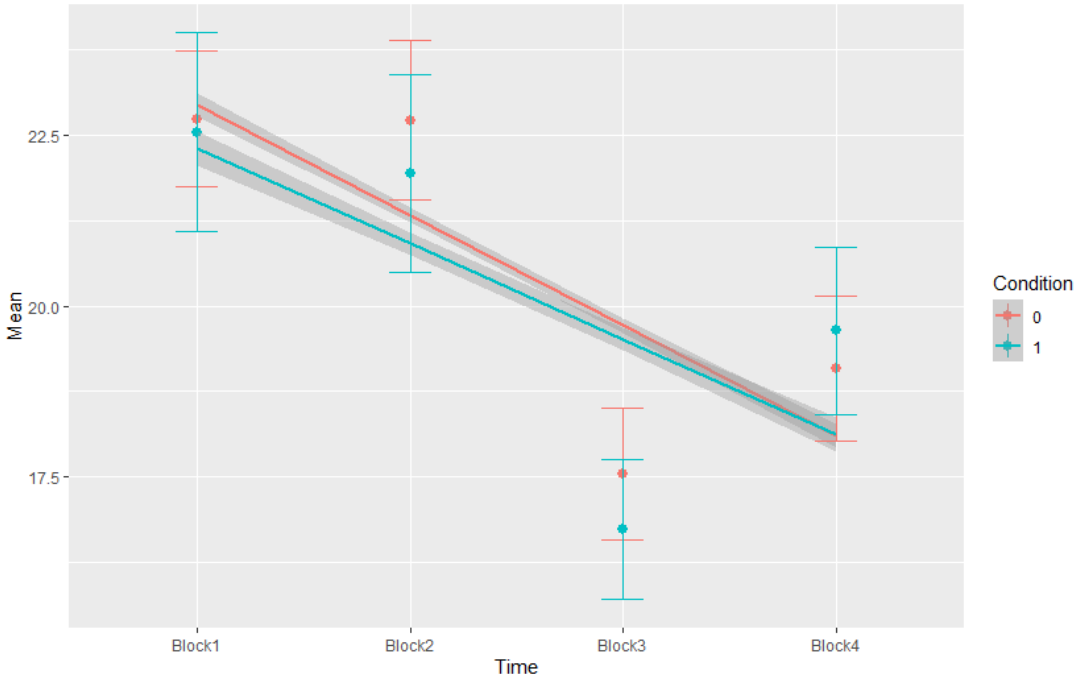


Figure B8:
Performance Plot by 4 Sets



[8] Note that lower performance = closer estimates to the stock price and more effective performance

Figure B9:
Attribution Plot by 21 Blocks

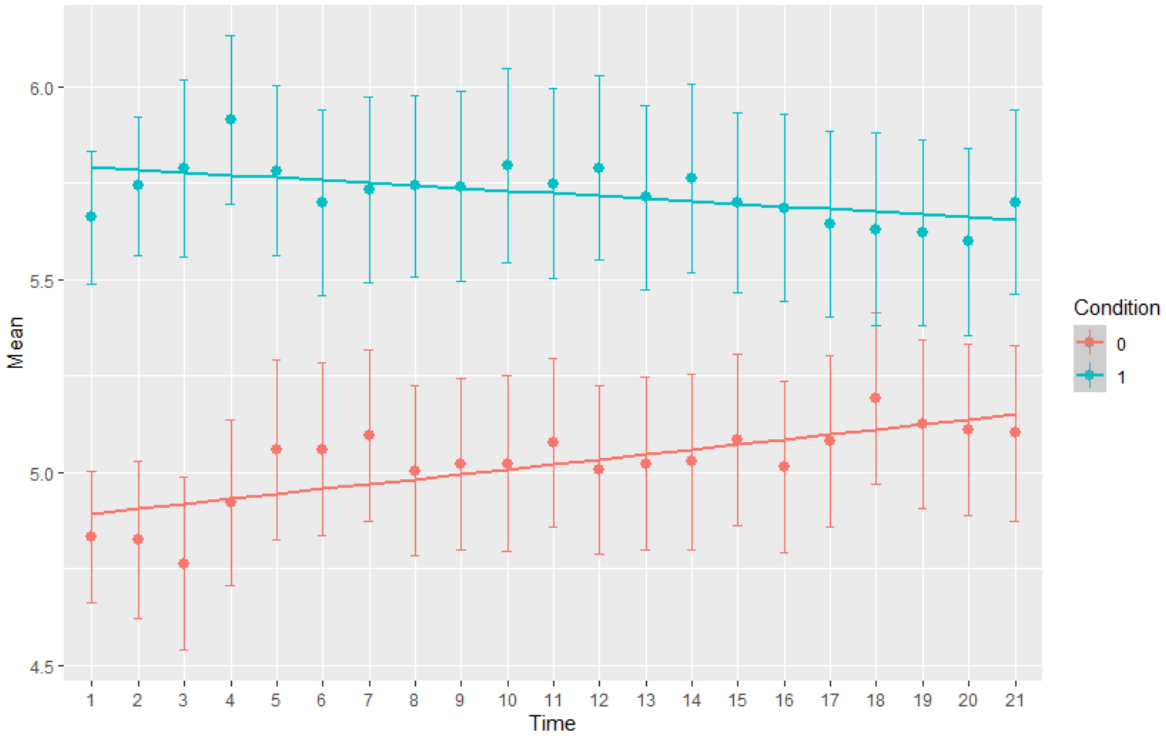


Figure B10:
Attribution Plot by 4 Sets

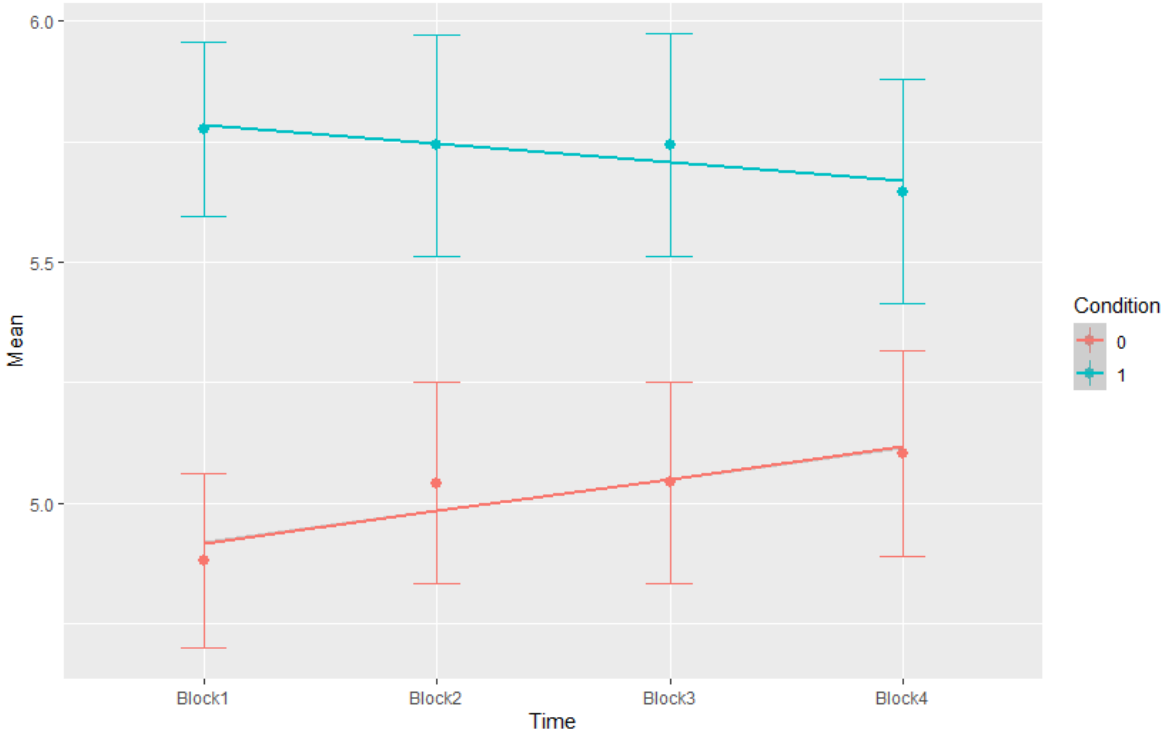


Figure B11:
Feedback Usefulness Plot by 21 Blocks

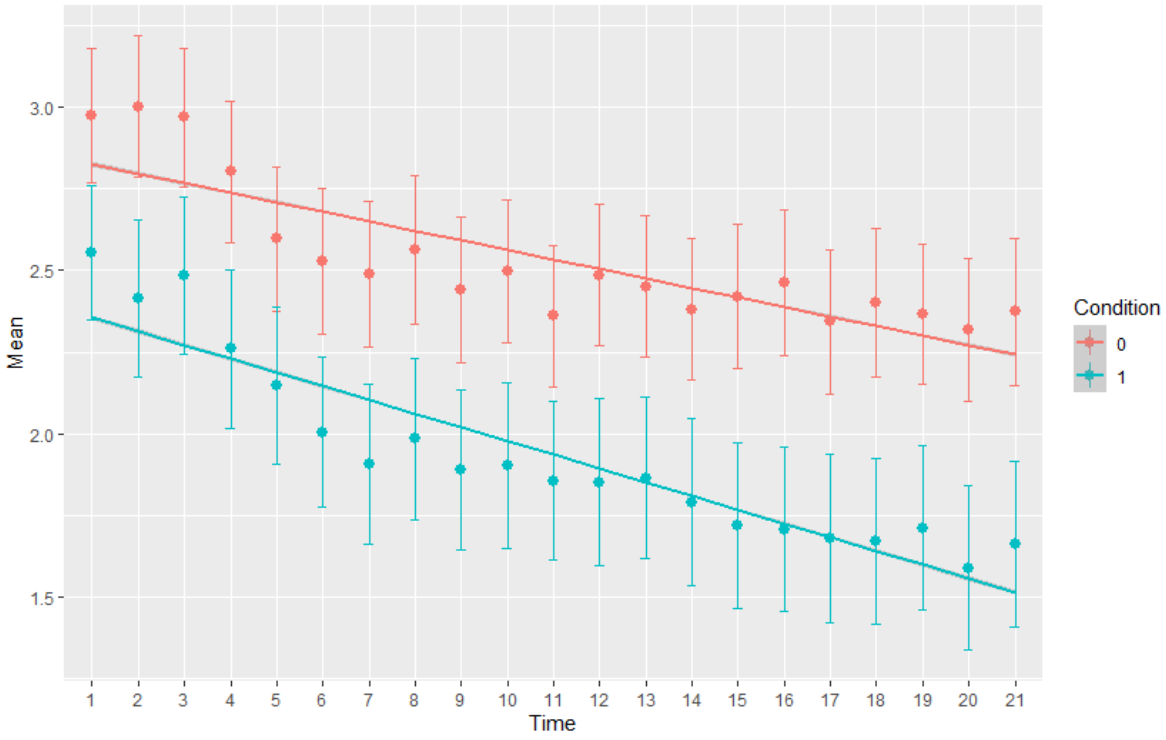


Figure B12:
Feedback Usefulness Plot by 4 Sets

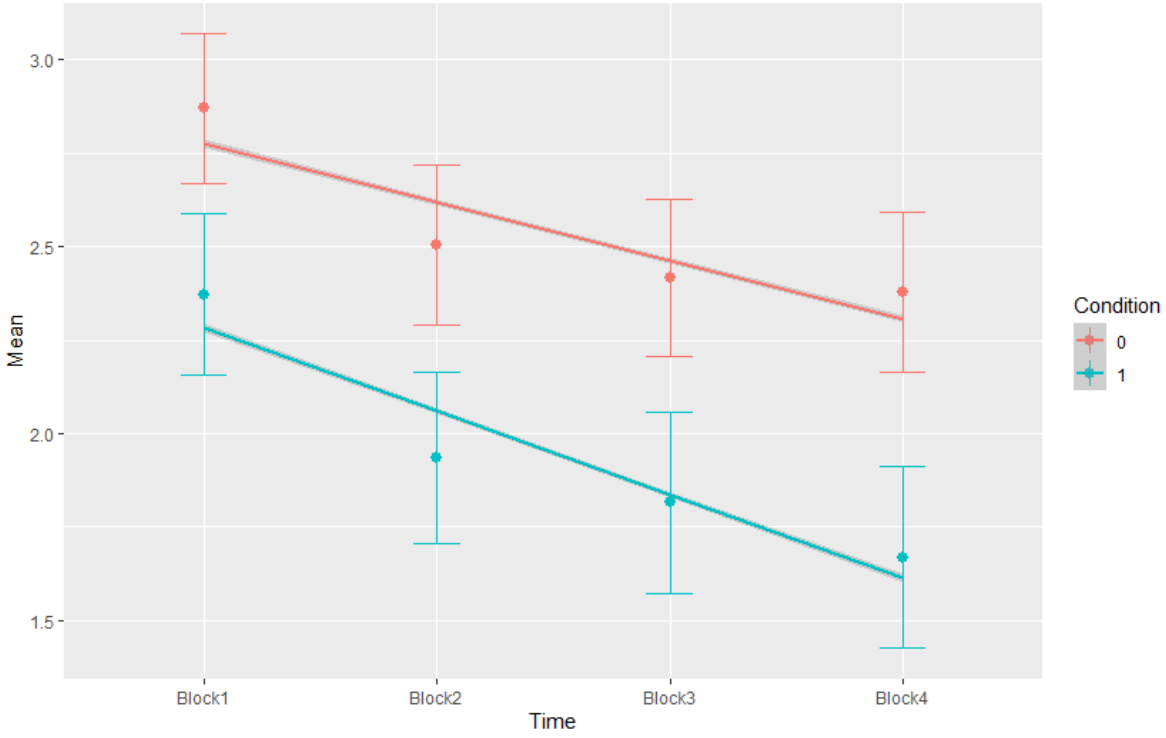


Figure B13:
Feedback Reactions Plot by 21 Blocks

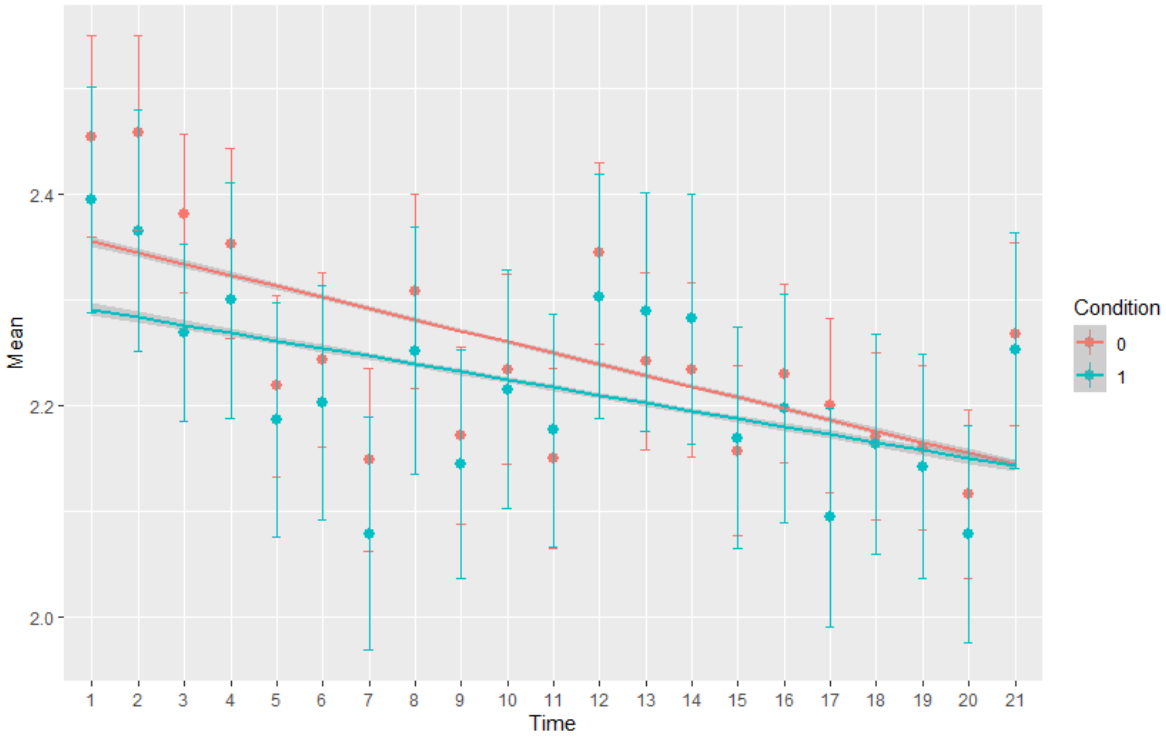


Figure B14:
Feedback Reactions Plot by 4 Sets

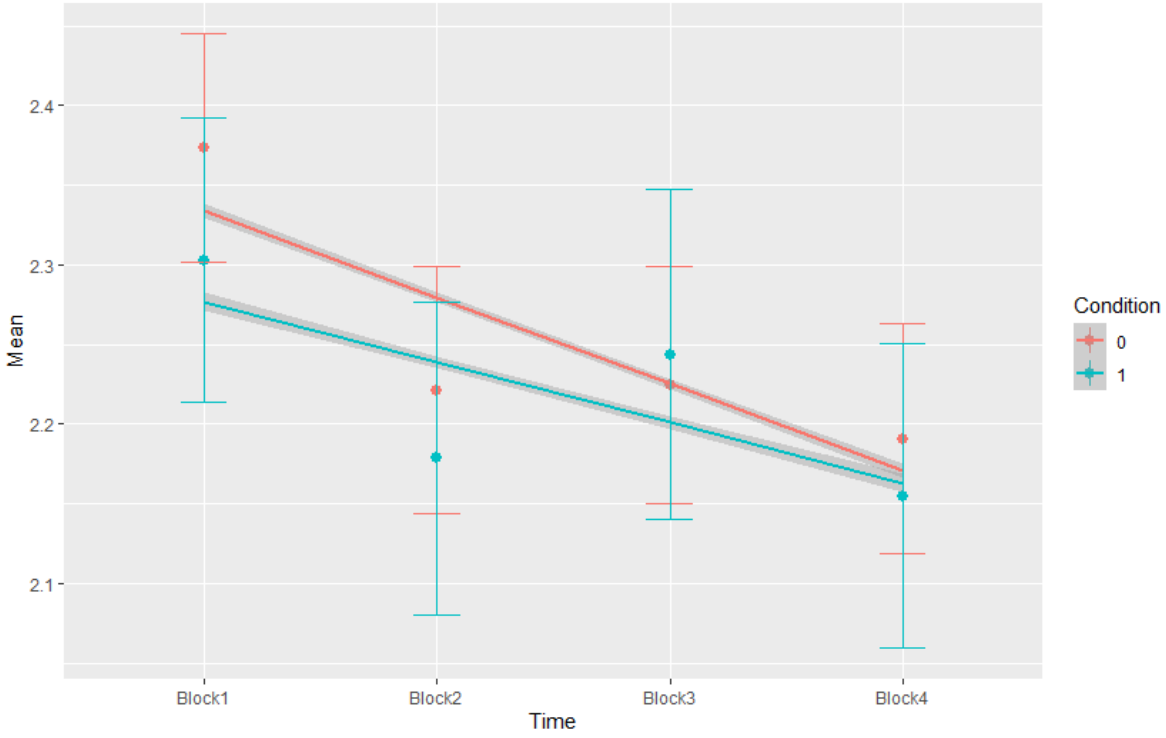


Figure B15:
Goal Commitment Plot by 21 Blocks

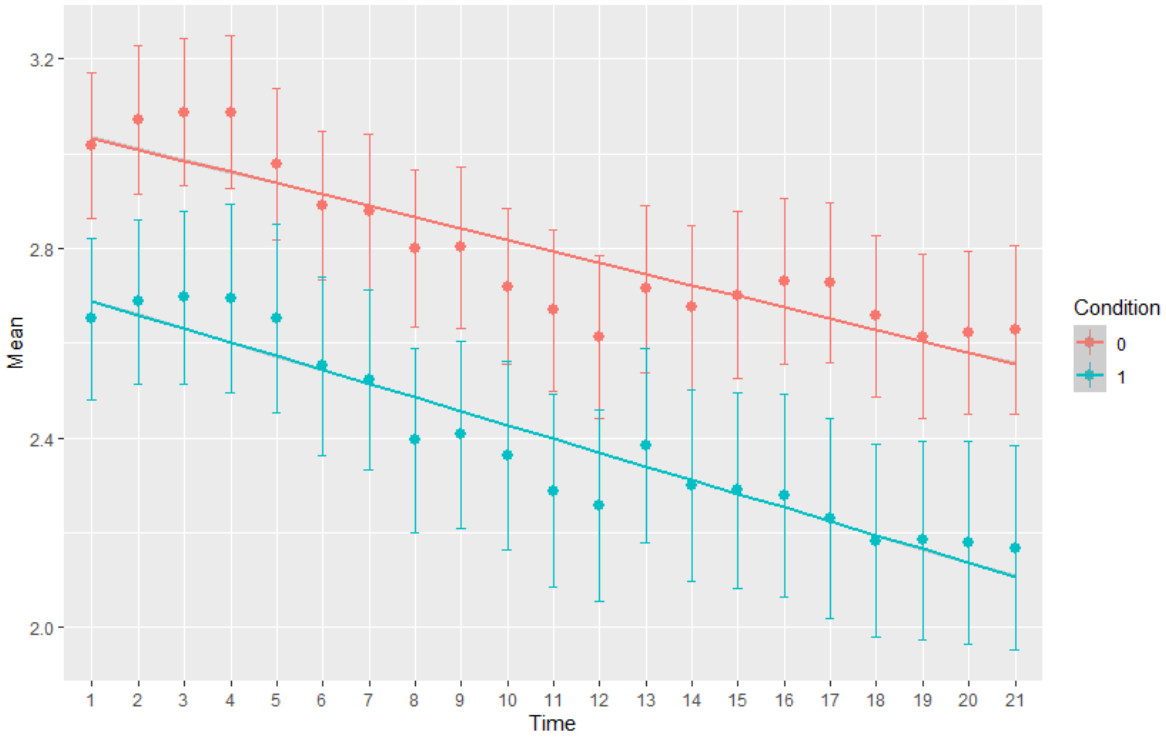


Figure B16:
Goal Commitment Plot by 4 Sets

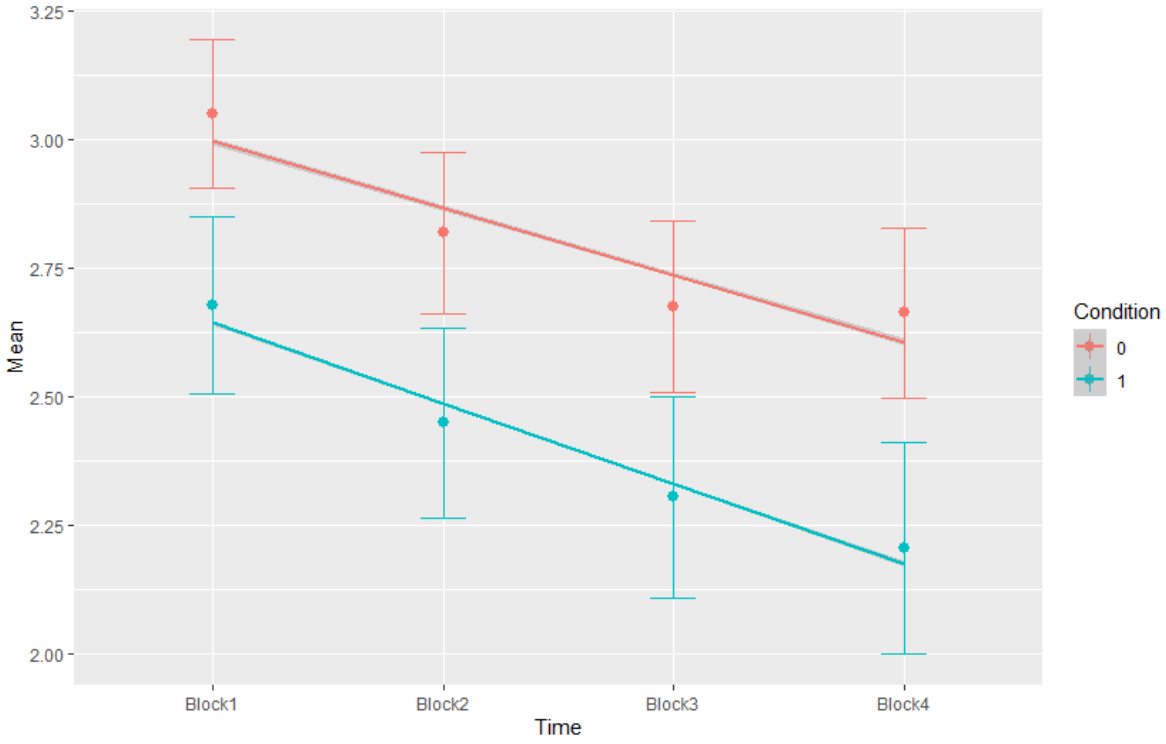


Figure B17:
Self-efficacy Plot by 21 Blocks

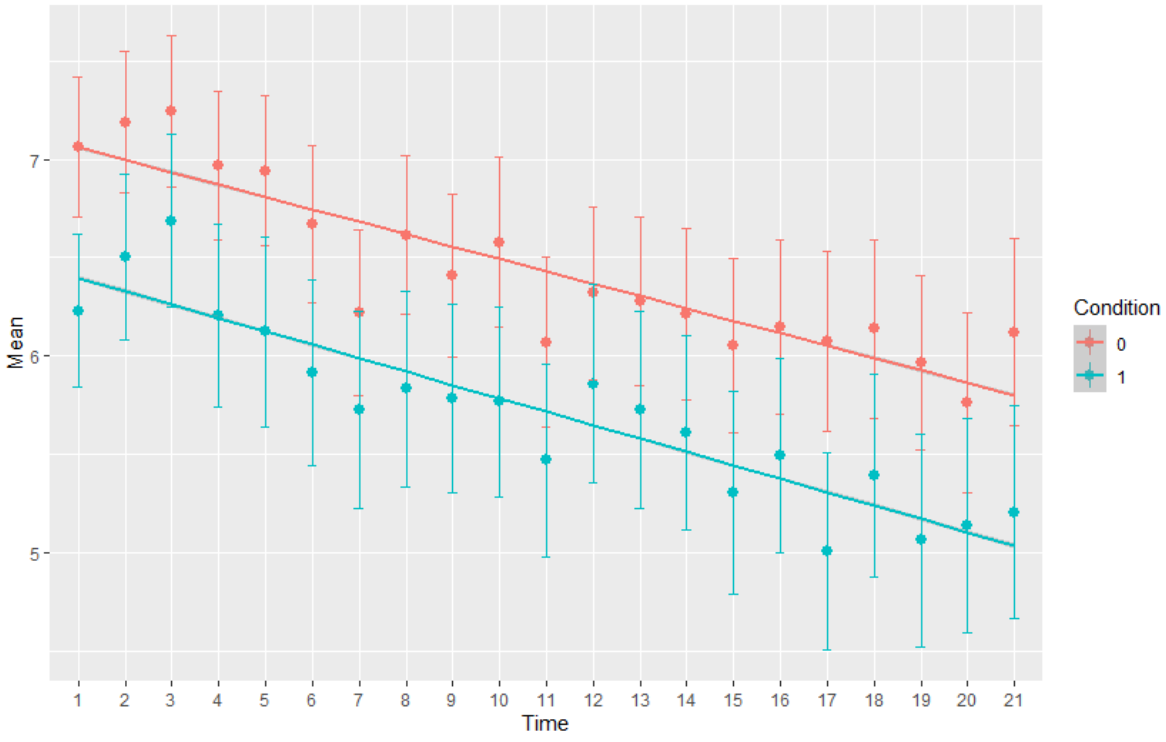
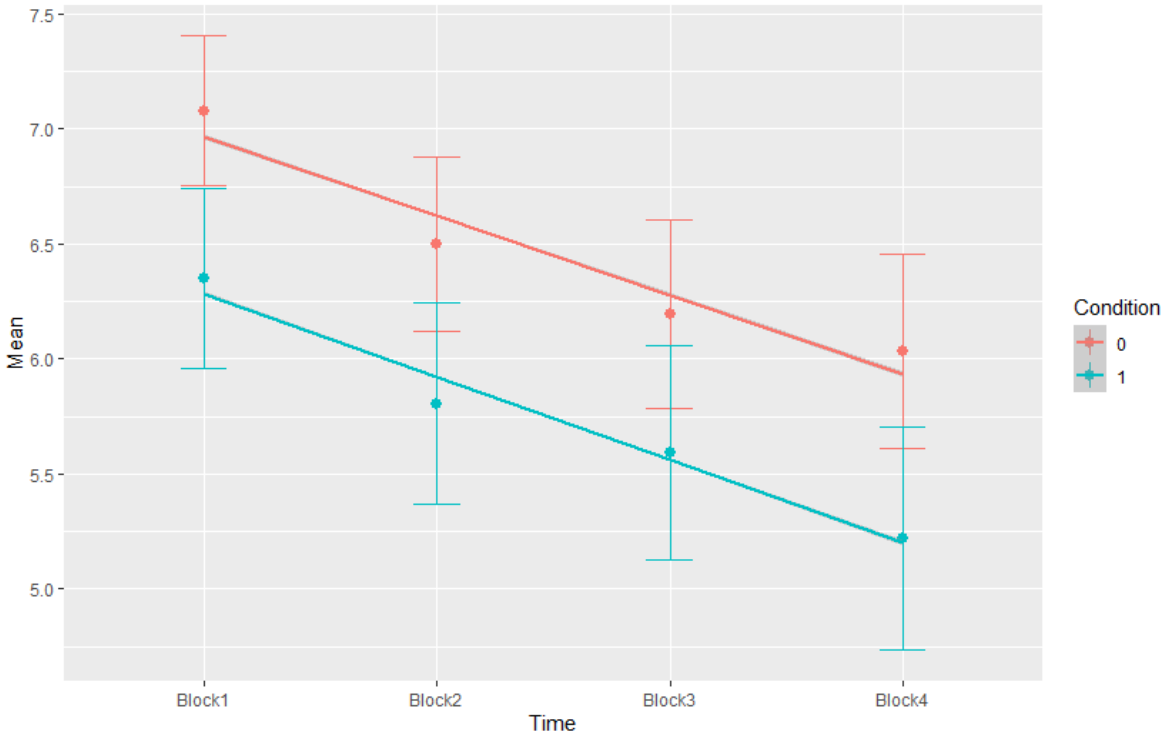


Figure B18:
Self-efficacy Plot by 4 Sets



APPENDIX C:
Descriptive and Correlation Tables

Table C1:
Variable Descriptive Statistics[9]

Variable	EXP Mean	CTRL Mean	EXP SD	CTRL SD	EXP SE	CTRL SE	Cohen's *d*	Cohen's *d* Text
E.1	7.06	6.23	2.43	2.14	0.18	0.19	-0.36	Small
E.2	7.19	6.5	2.46	2.33	0.18	0.21	-0.28	Small
E.3	7.24	6.69	2.62	2.44	0.19	0.22	-0.22	Small
E.4	6.97	6.21	2.6	2.58	0.19	0.23	-0.29	Small
E.5	6.94	6.12	2.61	2.67	0.19	0.24	-0.31	Small
E.6	6.67	5.92	2.73	2.62	0.2	0.24	-0.28	Small
E.7	6.22	5.73	2.86	2.78	0.21	0.25	-0.17	
E.8	6.62	5.83	2.76	2.75	0.2	0.25	-0.28	Small
E.9	6.41	5.79	2.83	2.67	0.21	0.24	-0.23	Small
E.10	6.58	5.77	2.95	2.68	0.22	0.24	-0.28	Small
E.11	6.07	5.47	2.94	2.73	0.22	0.25	-0.21	Small
E.12	6.32	5.86	3.01	2.8	0.22	0.25	-0.16	
E.13	6.28	5.73	2.93	2.79	0.22	0.25	-0.19	
E.14	6.22	5.61	2.96	2.75	0.22	0.25	-0.21	Small
E.15	6.05	5.31	3.03	2.86	0.22	0.26	-0.25	Small
E.16	6.15	5.5	3.03	2.75	0.22	0.25	-0.22	Small
E.17	6.08	5.01	3.13	2.79	0.23	0.25	-0.36	Small
E.18	6.14	5.4	3.12	2.86	0.23	0.26	-0.25	Small
E.19	5.97	5.07	3.01	2.99	0.22	0.27	-0.3	Small
E.20	5.76	5.14	3.1	3.01	0.23	0.27	-0.2	Small
E.21	6.12	5.21	3.25	3	0.24	0.27	-0.29	Small
G.1	24.96	25.79	18.8	23.58	1.39	2.14	0.04	
G.2	28.63	27.88	18.44	19.35	1.37	1.76	-0.04	
G.3	28.59	25.46	17.19	15.54	1.27	1.41	-0.19	
G.4	26.19	23.54	17.66	18.21	1.31	1.66	-0.15	
G.5	26.48	23.07	18.36	16.37	1.36	1.49	-0.19	
G.6	28.01	24.58	18.66	16.84	1.38	1.53	-0.19	
G.7	27.27	24.69	16.7	14.58	1.24	1.33	-0.16	
G.8	29.88	26.4	18.68	18.58	1.38	1.69	-0.19	
G.9	27.1	23.45	18.18	13.36	1.35	1.21	-0.22	Small
G.10	28.23	25	20.43	15.61	1.51	1.42	-0.17	
G.11	27.45	24.99	18.02	15.14	1.34	1.38	-0.15	
G.12	27.49	25.75	18.61	17.89	1.38	1.63	-0.1	
G.13	26.12	23.22	18.58	16.66	1.38	1.51	-0.16	
G.14	25.05	24.45	17.26	22.71	1.28	2.06	-0.03	
G.15	24.5	21.88	17.33	17.92	1.28	1.63	-0.15	
G.16	26.4	23.73	18.81	19.03	1.39	1.73	-0.14	

[9] E = Effort; G = Goal; P = Performance; A = Attributions; C = Goal Commitment; F = Feedback Usefulness; FR = Feedback Reactions; S = Self-efficacy; EXP = Experimental group; CTRL = Control group. G & P coded such that lower numbers indicate better performance/goals.

Table C1 (cont'd)

Variable	EXP Mean	CTRL Mean	EXP SD	CTRL SD	EXP SE	CTRL SE	Cohen's *d*	Cohen's *d* Text
G.17	24.94	22.52	18.38	15.79	1.36	1.44	-0.14	
G.18	25.57	22.81	16.84	15.78	1.25	1.43	-0.17	
G.19	24.43	23.72	17.06	20.34	1.26	1.85	-0.04	
G.20	25.42	23.05	20.73	16.87	1.54	1.53	-0.12	
G.21	25.52	22.64	20.69	15.72	1.53	1.43	-0.15	
P.1	24.77	25.63	12.24	16.83	0.91	1.53	0.06	
P.2	23.12	23.31	8.26	15.15	0.61	1.39	0.02	
P.3	17.15	17.97	8.54	14.16	0.63	1.29	0.07	
P.4	22.33	20.52	9.67	8.46	0.72	0.77	-0.2	
P.5	26.22	24.9	9.29	8.52	0.69	0.77	-0.15	
P.6	22.09	21.57	8.44	9.55	0.63	0.87	-0.06	
P.7	29.17	28.57	12.87	16.93	0.95	1.54	-0.04	
P.8	19.66	18.67	8.84	7.63	0.66	0.69	-0.12	
P.9	23.19	21.32	10.08	7.85	0.75	0.71	-0.2	Small
P.10	19.41	19.52	9.12	18.29	0.68	1.66	0.01	
P.11	20.7	20.12	8.45	7.86	0.63	0.71	-0.07	
P.12	13.31	13.56	7.74	6.63	0.58	0.6	0.04	
P.13	15.39	14.93	7.4	7.84	0.55	0.71	-0.06	
P.14	15.94	14.21	7.39	5.41	0.55	0.49	-0.26	Small
P.15	22.24	20.86	11.05	10.32	0.82	0.94	-0.13	
P.16	17.59	15.91	10.72	7.19	0.79	0.65	-0.18	
P.17	19.29	19.64	9.24	8.4	0.68	0.76	0.04	
P.18	17.64	16.78	8.06	9.35	0.6	0.85	-0.1	
P.19	17.89	17.7	8	8.37	0.59	0.76	-0.02	
P.20	22.44	21.5	10.3	10.18	0.77	0.93	-0.09	
P.21	16.92	17.26	12.22	9.3	1	1.32	0.03	
A.1	4.83	5.66	1.16	0.96	0.09	0.09	0.77	Medium
A.2	4.82	5.74	1.39	0.99	0.1	0.09	0.74	Medium
A.3	4.76	5.79	1.54	1.28	0.11	0.12	0.71	Medium
A.4	4.92	5.91	1.47	1.21	0.11	0.11	0.73	Medium
A.5	5.06	5.78	1.59	1.23	0.12	0.11	0.5	Small
A.6	5.06	5.7	1.54	1.33	0.11	0.12	0.44	Small
A.7	5.09	5.73	1.53	1.34	0.11	0.12	0.44	Small
A.8	5	5.74	1.51	1.3	0.11	0.12	0.52	Medium
A.9	5.02	5.74	1.52	1.37	0.11	0.12	0.49	Small
A.10	5.02	5.8	1.56	1.39	0.12	0.13	0.52	Medium
A.11	5.08	5.75	1.48	1.37	0.11	0.12	0.47	Small
A.12	5.01	5.79	1.49	1.32	0.11	0.12	0.55	Medium
A.13	5.02	5.71	1.53	1.33	0.11	0.12	0.48	Small
A.14	5.03	5.76	1.55	1.36	0.12	0.12	0.5	Small
A.15	5.08	5.7	1.51	1.31	0.11	0.12	0.43	Small

Table C1 (cont'd)

Variable	EXP Mean	CTRL Mean	EXP SD	CTRL SD	EXP SE	CTRL SE	Cohen's *d*	Cohen's *d* Text
A.16	5.01	5.69	1.52	1.35	0.11	0.12	0.46	Small
A.17	5.08	5.64	1.52	1.34	0.11	0.12	0.39	Small
A.18	5.19	5.63	1.52	1.38	0.11	0.13	0.3	Small
A.19	5.12	5.62	1.49	1.34	0.11	0.12	0.35	Small
A.20	5.11	5.6	1.53	1.35	0.11	0.12	0.34	Small
A.21	5.1	5.7	1.55	1.33	0.11	0.12	0.41	Small
C.1	3.02	2.65	1.06	0.95	0.08	0.09	-0.36	Small
C.2	3.07	2.69	1.07	0.97	0.08	0.09	-0.37	Small
C.3	3.09	2.7	1.06	1.02	0.08	0.09	-0.38	Small
C.4	3.09	2.69	1.1	1.1	0.08	0.1	-0.36	Small
C.5	2.98	2.65	1.1	1.11	0.08	0.1	-0.3	Small
C.6	2.89	2.55	1.07	1.05	0.08	0.1	-0.32	Small
C.7	2.88	2.52	1.12	1.06	0.08	0.1	-0.32	Small
C.8	2.8	2.4	1.13	1.08	0.08	0.1	-0.36	Small
C.9	2.8	2.41	1.16	1.09	0.09	0.1	-0.35	Small
C.10	2.72	2.36	1.13	1.11	0.08	0.1	-0.32	Small
C.11	2.67	2.29	1.16	1.13	0.09	0.1	-0.33	Small
C.12	2.61	2.26	1.17	1.12	0.09	0.1	-0.31	Small
C.13	2.72	2.38	1.21	1.14	0.09	0.1	-0.28	Small
C.14	2.68	2.3	1.19	1.13	0.09	0.1	-0.32	Small
C.15	2.7	2.29	1.2	1.15	0.09	0.1	-0.35	Small
C.16	2.73	2.28	1.19	1.19	0.09	0.11	-0.38	Small
C.17	2.73	2.23	1.15	1.18	0.09	0.11	-0.43	Small
C.18	2.66	2.18	1.17	1.13	0.09	0.1	-0.41	Small
C.19	2.61	2.18	1.18	1.17	0.09	0.11	-0.37	Small
C.20	2.62	2.18	1.17	1.19	0.09	0.11	-0.38	Small
C.21	2.63	2.17	1.22	1.2	0.09	0.11	-0.38	Small
F.1	2.98	2.55	1.4	1.15	0.1	0.1	-0.32	Small
F.2	3	2.41	1.48	1.33	0.11	0.12	-0.42	Small
F.3	2.97	2.48	1.45	1.34	0.11	0.12	-0.35	Small
F.4	2.8	2.26	1.49	1.35	0.11	0.12	-0.38	Small
F.5	2.6	2.15	1.51	1.34	0.11	0.12	-0.31	Small
F.6	2.53	2	1.52	1.27	0.11	0.12	-0.37	Small
F.7	2.49	1.9	1.53	1.36	0.11	0.12	-0.4	Small
F.8	2.56	1.98	1.57	1.37	0.12	0.12	-0.39	Small
F.9	2.44	1.89	1.53	1.36	0.11	0.12	-0.38	Small
F.10	2.5	1.9	1.5	1.41	0.11	0.13	-0.41	Small
F.11	2.36	1.86	1.49	1.36	0.11	0.12	-0.35	Small
F.12	2.49	1.85	1.49	1.43	0.11	0.13	-0.43	Small
F.13	2.45	1.86	1.48	1.37	0.11	0.12	-0.41	Small
F.14	2.38	1.79	1.47	1.43	0.11	0.13	-0.41	Small

Table C1 (cont'd)

Variable	EXP Mean	CTRL Mean	EXP SD	CTRL SD	EXP SE	CTRL SE	Cohen's *d*	Cohen's *d* Text
F.15	2.42	1.72	1.51	1.41	0.11	0.13	-0.48	Small
F.16	2.46	1.71	1.53	1.4	0.11	0.13	-0.51	Medium
F.17	2.34	1.68	1.52	1.43	0.11	0.13	-0.45	Small
F.18	2.4	1.67	1.55	1.41	0.11	0.13	-0.49	Small
F.19	2.37	1.71	1.48	1.4	0.11	0.13	-0.45	Small
F.20	2.32	1.59	1.5	1.4	0.11	0.13	-0.5	Medium
F.21	2.37	1.66	1.53	1.41	0.11	0.13	-0.48	Small
FR.1	2.45	2.39	0.65	0.59	0.05	0.05	-0.1	
FR.2	2.46	2.37	0.63	0.63	0.05	0.06	-0.15	
FR.3	2.38	2.27	0.51	0.47	0.04	0.04	-0.23	Small
FR.4	2.35	2.3	0.62	0.62	0.05	0.06	-0.09	
FR.5	2.22	2.19	0.59	0.62	0.04	0.06	-0.05	
FR.6	2.24	2.2	0.56	0.62	0.04	0.06	-0.07	
FR.7	2.15	2.08	0.59	0.61	0.04	0.06	-0.12	
FR.8	2.31	2.25	0.63	0.65	0.05	0.06	-0.09	
FR.9	2.17	2.14	0.57	0.6	0.04	0.05	-0.05	
FR.10	2.23	2.22	0.61	0.63	0.05	0.06	-0.03	
FR.11	2.15	2.18	0.58	0.61	0.04	0.06	0.04	
FR.12	2.34	2.3	0.59	0.64	0.04	0.06	-0.07	
FR.13	2.24	2.29	0.57	0.63	0.04	0.06	0.08	
FR.14	2.23	2.28	0.56	0.66	0.04	0.06	0.08	
FR.15	2.16	2.17	0.55	0.58	0.04	0.05	0.02	
FR.16	2.23	2.2	0.58	0.6	0.04	0.05	-0.06	
FR.17	2.2	2.09	0.56	0.57	0.04	0.05	-0.19	
FR.18	2.17	2.16	0.54	0.58	0.04	0.05	-0.01	
FR.19	2.16	2.14	0.53	0.59	0.04	0.05	-0.03	
FR.20	2.12	2.08	0.55	0.57	0.04	0.05	-0.07	
FR.21	2.27	2.25	0.59	0.62	0.04	0.06	-0.03	
S.1	5.25	4.71	1.81	1.79	0.13	0.16	-0.3	Small
S.2	5.81	5.38	1.9	1.98	0.14	0.18	-0.22	Small
S.3	5.99	5.45	1.71	1.97	0.13	0.18	-0.3	Small
S.4	6.32	5.8	1.77	2.04	0.13	0.19	-0.27	Small
S.5	6.29	5.9	1.8	2	0.13	0.18	-0.21	Small
S.6	6.11	5.52	1.79	2.07	0.13	0.19	-0.31	Small
S.7	6.08	5.66	1.89	1.98	0.14	0.18	-0.22	Small
S.8	6.07	5.62	1.97	1.99	0.15	0.18	-0.23	Small
S.9	6.21	5.72	1.95	2.05	0.14	0.19	-0.25	Small
S.10	6.15	5.72	1.94	2.12	0.14	0.19	-0.21	Small
S.11	6.09	5.76	1.97	2.11	0.15	0.19	-0.16	
S.12	6.19	5.79	2	2.1	0.15	0.19	-0.2	
S.13	6.26	6.04	2.07	2.14	0.15	0.19	-0.11	

Table C1 (cont'd)

Variable	EXP Mean	CTRL Mean	EXP SD	CTRL SD	EXP SE	CTRL SE	Cohen's *d*	Cohen's *d* Text
S.14	6.28	6.25	2.13	2.17	0.16	0.2	-0.01	
S.15	6.3	6.23	2.17	2.2	0.16	0.2	-0.03	
S.16	6.24	6.05	2.05	2.18	0.15	0.2	-0.09	
S.17	6.21	6.12	2.08	2.25	0.15	0.2	-0.04	
S.18	6.28	6.13	2.07	2.22	0.15	0.2	-0.07	
S.19	6.29	6.28	2.04	2.22	0.15	0.2	0	
S.20	6.42	6.27	2.1	2.22	0.16	0.2	-0.07	
S.21	6.25	6.31	2.09	2.26	0.16	0.21	0.03	
E.Set1	7.08	6.35	2.23	2.16	0.17	0.2	-0.33	Small
E.Set2	6.5	5.81	2.59	2.42	0.19	0.22	-0.28	Small
E.Set3	6.19	5.6	2.8	2.59	0.21	0.24	-0.22	Small
E.Set4	6.03	5.22	2.89	2.7	0.21	0.25	-0.29	Small
G.Set1	26.97	25.15	15.49	15.31	1.15	1.39	-0.12	
G.Set2	28.1	24.82	16.77	13.83	1.24	1.26	-0.21	Small
G.Set3	26.12	24.06	16.17	15.09	1.2	1.37	-0.13	
G.Set4	25.38	23.08	17.02	15.14	1.26	1.38	-0.14	
P.Set1	22.73	22.54	6.8	8.05	0.5	0.73	-0.03	
P.Set2	22.72	21.93	7.97	8.04	0.59	0.73	-0.1	
P.Set3	17.54	16.74	6.62	5.72	0.49	0.52	-0.13	
P.Set4	18.69	18.26	7.37	6.56	0.55	0.6	-0.06	
A.Set1	4.88	5.78	1.25	1	0.09	0.09	0.78	Medium
A.Set2	5.04	5.74	1.43	1.28	0.11	0.12	0.51	Medium
A.Set3	5.04	5.74	1.43	1.28	0.11	0.12	0.51	Medium
A.Set4	5.1	5.65	1.46	1.3	0.11	0.12	0.39	Small
C.Set1	3.05	2.68	0.99	0.96	0.07	0.09	-0.38	Small
C.Set2	2.82	2.45	1.07	1.03	0.08	0.09	-0.35	Small
C.Set3	2.68	2.3	1.14	1.09	0.08	0.1	-0.33	Small
C.Set4	2.66	2.2	1.13	1.14	0.08	0.1	-0.4	Small
F.Set1	2.87	2.37	1.36	1.2	0.1	0.11	-0.38	Small
F.Set2	2.5	1.94	1.46	1.27	0.11	0.12	-0.41	Small
F.Set3	2.41	1.82	1.43	1.34	0.11	0.12	-0.43	Small
F.Set4	2.38	1.67	1.46	1.35	0.11	0.12	-0.5	Medium
FR.Set1	2.37	2.3	0.49	0.5	0.04	0.05	-0.14	
FR.Set2	2.22	2.18	0.53	0.55	0.04	0.05	-0.08	
FR.Set3	2.22	2.24	0.51	0.57	0.04	0.05	0.04	
FR.Set4	2.19	2.15	0.49	0.53	0.04	0.05	-0.07	
S.Set1	5.93	5.45	1.59	1.78	0.12	0.16	-0.29	Small
S.Set2	6.12	5.65	1.82	1.97	0.14	0.18	-0.25	Small
S.Set3	6.23	6.01	1.99	2.07	0.15	0.19	-0.1	
S.Set4	6.28	6.19	2	2.17	0.15	0.2	-0.04	

Table C2:
Correlation Table of Study Variables

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Condition	0.40	0.49								
2. Effort	6.15	2.42	-.14*							
3. Goal Level	25.64	14.48	-.08	-.22**						
4. Performance	20.14	6.13	-.04	-.15**	.52**					
5. Performance Attributions	5.30	1.28	.27**	.03	-.07	-.07				
6. Goal Commitment	2.64	1.03	-.19**	.66**	-.19**	-.19**	-.04			
7. Feedback Usefulness	2.29	1.34	-.22**	.49**	-.18**	-.18**	-.13*	.56**		
8. Feedback Reactions	2.24	0.48	-.03	.39**	-.23**	-.27**	.01	.44**	.62**	
9. Self-efficacy	6.03	1.81	-.08	.19**	-.37**	-.28**	-.05	.36**	.35**	.31**

Table C2 (cont'd)

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
10. Trait Learning Orientation	4.13	0.52	-.06	.13*	-.08	.04	.08	.18**	.16**	.16**	.11									
11. Trait Prove Orientation	3.46	0.64	-.01	.08	.05	-.04	-.11	.05	.04	-.04	-.01	.17**								
12. Trait Avoid Orientation	2.97	0.80	.04	-.02	.11*	.02	-.02	-.12*	-.06	-.10	-.10	-.31**	.28**							
13. Trait Positive Affect	3.53	0.64	-.05	.08	-.03	.08	-.01	.19**	.14*	.22**	.09	.39**	.07	-.22**						
14. Trait Negative Affect	2.25	0.62	-.03	-.02	.09	.00	.01	-.06	.04	-.08	.02	-.12*	.14*	.25**	-.17**					
15. State Learning Orientation	3.59	0.62	.01	.46**	-.27**	-.16**	.11*	.60**	.37**	.39**	.24**	.43**	.05	-.16**	.26**	-.06				
16. State Prove Orientation	3.03	0.63	-.04	.33**	-.04	-.04	-.04	.43**	.27**	.15**	.08	.12*	.46**	.23**	.09	.13*	.47**			
17. State Avoid Orientation	2.78	0.65	-.07	.17**	.15**	.04	-.11*	.18**	.12*	-.07	-.03	-.18**	.39**	.48**	-.09	.27**	.14*	.68**		
18. State Positive Affect	4.16	2.50	-.19**	.57**	-.15**	-.18**	-.11	.63**	.64**	.63**	.28**	.24**	.09	-.12*	.25**	.05	.53**	.38**	.11	
19. State Negative Affect	2.70	2.11	-.23**	.27**	.17**	.10	-.14*	.26**	.32**	-.05	.12*	.08	.21**	.08	-.01	.35**	.10	.32**	.34**	.42**

APPENDIX D:
Supplementary Materials

Appendix D1:

Sample R Code for lavaan Growth Curve Modeling

```
model =
 "


i =~ 1*E.1 + 1*E.2 + 1*E.3 + 1*E.4 + 1*E.5 + 1*E.6 + 1*E.7 + 1*E.8 + 1*E.9 + 1*E.10 +
1*E.11 +

1*E.12 + 1*E.13 + 1*E.14 + 1*E.15 + 1*E.16 + 1*E.17 + 1*E.18 + 1*E.19 + 1*E.20 + 1*E.21


s =~ 0*E.1 + 1*E.2 + 2*E.3 + 3*E.4 + 4*E.5 + 5*E.6 + 6*E.7 + 7*E.8 + 8*E.9 + 9*E.10 +
10*E.11 +

11*E.12 + 12*E.13 + 13*E.14 + 14*E.15 + 15*E.16 + 16*E.17 + 17*E.18 + 18*E.19 + 19*E.20
+ 20*E.21



i ~ Condition

s ~ Condition


"
```

REFERENCES

REFERENCES

Alderfer, C. (1987). An intergroup perspective on group dynamics. In J.W. Lorsch (Ed.), *Handbook of organizational behavior* (pp.190-222). Englewood Cliffs, NJ: Prentice-Hall.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*(2), 88-93.

Avery, D.R. (2003). Reactions to diversity in recruitment advertising—Are differences black and white? *Journal of Applied Psychology, 88*(4), 672-679.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Ballard, T., Yeo, G., Neal, A., & Farrell, S. (2016). Departures from optimality when pursuing multiple approach or avoidance goals. *Journal of Applied Psychology, 101*(7), 1056-1066.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50*(2), 248-287.

Bandura, A. 1997). *Self-efficacy: The exercise of control.* New York, NY: W.H. Freeman.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*, 9-44.

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes, 38*(1), 92-113.

Bandura, A., & Locke, E.A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87-99.

Bauer, G.R., Braimoh, J., Scheim, A.I., & Dharma, C. (2017). Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLoS One, 12*(5), e0178043.

Beck, J.W., & Schmidt, A.M. (2012). Taken out of context? Cross-level effects of between-person self-efficacy and difficulty on the within-person relationship of self-efficacy with resource allocation and performance. *Organizational Behavior and Human Decision Processes, 119,* 195-208.

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.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*(2), 296-316.

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

Bergeron, D.M., Block, C.J., & Echtenkamp, A. (2006). Disabling the able: Stereotype threat and women's math performance. *Human Performance, 19*(2), 133-158.

Bliese, P.D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K.J. Klein & S.W.J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349-381). San Francisco: Jossey-Bass.

Brett, J.F., & Atwater, L.E. (2001). 360° Feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930-942.

Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45*(1), 180-185.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Campbell, J.P., & Pritchard, R.D. (1976). Motivation theory in industrial and organizational psychology. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp.63-130). Chicago, IL: Rand McNally.

Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 47-74.

Carr, P.B., & Steele, C.M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853-859.

Carr, P.B., & Steele, C.M. (2010). Stereotype threat affects financial decision making. *Psychological Science, 21*(10), 1411-1416.

Carver, C.S., & Scheier, M.F. (1998). *On the self-regulation of behavior*. Cambridge, England: Cambridge University Press.

Casad, B.J., & Merritt, S.M. (2014). The importance of stereotype threat mechanisms in workplace outcomes. *Industrial and Organizational Psychology, 7*(3), 413-419.

Chalabaev, A., Dematte, E., Sarrazin, P., & Fontayne, P. (2014). Creating regulatory fit under stereotype threat: Effects on performance and self-determination among junior high school students. *Revue Internationale de Psychologie Sociale* [International Review of Social Psychology], *27*(3), 119-132.

Cheryan, S., & Bodenhausen, G. V. (2000). When positive stereotypes threaten intellectual performance: The psychological hazards of "model minority" status. *Psychological Science, 11*(5), 399-402.

Cohen, G.L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, implications. *Current Directions in Psychological Science, 17*(6), 365-369.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*(2), 331-335.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*, 608-630.

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*(6), 1639-1655.

Davies, P.G., Spencer, S.J., Quinn, D.M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*(12), 1615-1628.

Deemer, E. D., Smith, J. L., Carroll, A. N., & Carpenter, J. P. (2014). Academic procrastination in stem: Interactive effects of stereotype threat and achievement goals. *The Career Development Quarterly, 62*(2), 143-155.

Deemer, E. D., Thoman, D. B., Chase, J. P., & Smith, J. L. (2014). Feeling the threat: Stereotype threat as a contextual barrier to women's science career choice intentions. *Journal of Career Development, 41*(2), 141-158.

DeShon, R.P. (2012). Multivariate dynamics in organization science. In S.W.J. Kozlowski (Ed.), *The oxford handbook of organizational psychology* (Vol. 1, pp. 117-142). New York, NY: Oxford University Press.

Delisle, M.N., Guay, F., Senécal, C., & Larose, S. (2009). Predicting stereotype endorsement and academic motivation in women in science programs: A longitudinal model. *Learning and Individual Differences, 19*(4), 468-475.

Donovan, J.J., & Williams, K.J. (2003). Missing the mark: Effects of time and causal attributions on goal revision in response to goal-performance discrepancies. *Journal of Applied Psychology, 88*, 379-390.

Earley, P.C., Connolly, T., & Ekegren, G. (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology, 74*, 24-33.

Endres, M.L., Chowdhury, S., & Milner, M. (2009). Ambiguity tolerance and accurate assessment of self-efficacy in a complex decision task. *Journal of Management and Organization, 15*, 31-46.

Fisher, S.I., & Ford, J.K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.

Fogliati, V.J., & Bussey, K. (2013). Stereotype threat reduces motivation to improve: Effects of stereotype threat and feedback on women's intentions to improve mathematical ability. *Psychology of Women Quarterly, 37*(3), 310-324.

Forbes, C. E., Schmader, T., & Allen, J. J. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3*(3), 253-261.

Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692-731.

Grand, J. A. (2017). Brain drain? An examination of stereotype threat effects during training on knowledge acquisition and organizational effectiveness. *Journal of Applied Psychology, 102*(2), 115-150.

Grand, J. A., Golubovich, J., Ryan, A. M., & Schmitt, N. (2013). The detection and influence of problematic item content in ability tests: An examination of sensitivity review practices for personnel selection test development. *Organizational Behavior and Human Decision Processes, 121,* 158–173.

Grimm, L.R., Marman, A.B., Maddox, W.T., & Baldwin, G.C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology, 96*(2), 288-304.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*(3), 281-291.

Gully, S.M., Payne, S.C., Kiechel, K.L., & Whiteman, J.K. (1999, April). *Affective reactions and performance outcomes of error-based training*. Paper presented at the 14th Annual Society for Industrial & Organizational Psychology Conference, Atlanta, GA.

Hagger, M. S., Chatzisarantis, N. L., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., ... & Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546-573.

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

Heckhausen, J. (1991). *Motivation and action.* Berlin, Germany: Springer-Verlag.

Hirnstein, M., Freund, N., & Hausmann, M. (2015). Gender stereotyping enhance verbal fluency performance in men (and women*). Zeitschrift für Psychologie* [Journal of Psychology], *220*, 70-77.

Horvath, M., Scheu, C.R., & DeShon, R.P. (2001). *Goal orientation: Integrating theory and measurement.* Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Huber, M.E., Seitchik, A.E., Brown, A.J., Sternad, D., & Harkins, S.G. (2015). The effect of stereotype threat on performance of a rhythmic motor skill. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 525-541.

Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology, 49*(3), 550-565.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology, 90*(3), 453-467.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology, 93*(4), 544.

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

Jamieson, J.P., & Harkins, S.G. (2011). The intervening task method: Implications for measuring mediation. *Personality and Social Psychology Bulletin, 37*, 652-661.

Jamieson, J.P., & Harkins, S.G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes & Intergroup Relations, 15*(3), 291-304.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General, 137*(4), 691.

Kalokerinos, E.K., von Hippel, C., & Zacher, H. (2014). Is stereotype threat a useful construct for organizational psychology research and practice? *Industrial and Organizational Psychology, 7*, 381-402.

Kanfer, R. (1990). Motivation theory in I/O psychology. In M.D. Dunnette & L.M. Hough (eds.), *The handbook of industrial & organizational psychology* (2nd ed., pp.75-170). Pal Alto, CA: Consulting Psychologists Press.

Kanfer, R., & Ackerman, P.L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*(4), 657-690.

Kanfer, R., Frese, M., & Johnson, R. E. (2017). Motivation related to work: A century of progress. *Journal of Applied Psychology, 102,* 338-355.

Kanfer, R., & Heggestad, E. D. (1999). Individual differences in motivation: Traits and self-regulatory skills. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 293–309). Washington, DC: American Psychological Association.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology, 90*(4), 677.

Keller, J. (2010). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology, 77*(2), 323-338.

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

Kelley, T.L. (1927). *Interpretation of educational measurements*. Oxford, England: World Book.

Kernan, M.C., & Lord, R.G. (1990). Effect of valence, expectancies, and goal-performance discrepancies in single and multiple goal environments. (1990). *Journal of Applied Psychology, 75*(2), 194-203.

Kinias, Z., & Sim, J. (2016). Facilitating women's success in business: Interrupting the process of stereotype threat through affirmation of personal values. *Journal of Applied Psychology, 101*(11), 1585-1597.

Kinzler, K.D., & DeJesus, J.M. (2013). Northern = smart and Southern = nice: The development of accent attitudes in the United States. *Quarterly Journal of Experimental Psychology, 66*(6), 1146-1158.

Klein, H. J., Cooper, J. T., Molloy, J. C., & Swanson, J. A. (2014). The assessment of commitment: Advantages of a unidimensional, target-free approach. *Journal of Applied Psychology, 99*(2), 222-238.

Klein, H. J., Cooper, J. T., & Monahan, C. A. (2013). Goal commitment. In E. A. Locke & G. P. Latham (Eds.), *New developments in goal setting and task performance* (pp. 65-89). New York, NY: Routledge.

Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.

Koy, A., & Yeo, G. (2008). BIS sensitivity, negative affect and performance: Dynamic and multilevel relationships. *Human Performance, 21*(2), 198-225.

Kozlowski, S.W.J., Toney, R.J., Mullins, M.E., Weissbein, D.A., Brown, K.G., & Bell, B.S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

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

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science, 19*, 168–175.

Kukla, A. (1972). Foundations of an attributional theory of performance. *Psychological Review, 79*, 454-470.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology, 1*(4), 379-392.

Latsch, M., & Hannover, B. (2014). Smart girls, dumb boys!? How the discourse on "failing boys" impacts performances and motivational goal orientation in German school students. *Social Psychology, 45*, 112-126.

Lin, S.-H., & Johnson, R.E. (2015). A suggestion to improve a day keeps your depletion away: Examining promotive and prohibitive voice behaviors within a regulatory focus and ego depletion framework. *Journal of Applied Psychology, 100*, 1381-1397.

Lindsley, D.H., Brass, D.J., & Thomas, J.B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review, 20*(3), 645-678.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*: Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705.

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34-50.

Mangels, J.A., Good, C., Whiteman, R.C., Maniscalco, B., & Dweck, C.S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience, 7*, 230-241.

Martell, R. F., Lane,D. M.,& Emrich, C. G. (1996). Male-female differences: A computer simulation. *American Psychologist, 38*, 379–387.

Martinko, M. J., Harvey, P., & Dasborough, M. T. (2011). Attribution theory in the organizational sciences: A case of unrealized potential. *Journal of Organizational Behavior, 32*(1), 144-149.

Maslyn, J.M., & Uhl-Bien, M. (2001). Leader-member exchange and its dimensions: Effects of self-effort and other's effort on relationship quality. *Journal of Applied Psychology, 86*(4), 697-708.

Masuda, A.L., Locke, E.A., & Williams, K.J. (2014). The effects of simultaneous learning and performance goals on performance: An inductive exploration. *Journal of Cognitive Psychology, 27*, 37-52.

McAuley, E., Duncan, T.E., & Russell, D.W. (1992). Measuring causal attributions: The revised causal dimension scale (CDSII). *Personality and Social Psychology Bulletin, 18*(5), 566-573.

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

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247.

Murphy, M.C., Steele, C.M., & Gross, J.J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18,* 879-885.

Nguyen, H.H.D., & Ryan, A.M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314-1334.

Nussbaum, A.D., & Steele, C.M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology, 43*, 127-134.

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

Osborne, J.W., & Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of colour might be the most likely to withdraw. *Educational Psychology, 26*(4), 563-577.

Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology, 19*(4), 338-356.

Park, G., Schmidt, A.M., Scheu, C., & DeShon, R.P. (2007). A process model of goal orientation and feedback seeking. *Human Performance, 20,* 119-145.

Pinder, C.C. (1976). Additivity versus nonadditivity of intrinsic and extrinsic incentives: Implications for work motivation, performance, and attitudes. *Journal of Applied Psychology, 61*, 693-700.

Ployhart, R.E., & Vandenberg, R.J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management, 36,* 94-120.

Powers, W. T. (1973). *Behavior: The control of perception*. Chicago: Aldine.

Powers, W. T. (1991). Commentary on bandura's "human agency." *American Psychologist, 46*(2), 151-153.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic groups* (College Board Report 93-1, ETS No. 94.27). New York: College Entrance Examination Board.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Roberson, L., Deitch, E. A., Brief, A. P., & Block, C. J. (2003). Stereotype threat and feedback seeking in the workplace. *Journal of Vocational Behavior, 62*(1), 176-188.

Ryan, A.M. & Nguyen, H.H.D. (2017). Publication bias and stereotype threat research: A reply to Zigerell. *Journal of Applied Psychology, 102*(8), 1169-1177.

Sackett, P. R., & Ryan, A. M. (2012). Concerns about generalizing stereotype threat research findings to operational high stakes testing. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 249–263). New York, NY: Oxford Press.

Saenz, D.S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition, 12*, 61-74.

Samuelson, H.L., Fernandez, J.R., & Grand, J.A. (2017). "Life doesn't happen at the between-person level," or a cautionary note on generating scientific inferences through meta-analyses. *Industrial and Organizational Psychology, 10*(3), 459-464.

Scherbaum, C.A., & Ferreter, J.M. (2008). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*(2), 347-367.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.

Schmader, T., & Beilock, S. (2012). An integration of processes that underlie stereotype threat. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 34-50). New York, NY: Oxford Press.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336.

Schmidt, F.L. (1988). The problem of group differences ability test scores in employment selection. *Journal of Vocational Behavior, 33*, 272-292.

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

Seitchik, A.E., & Harkins, S.G. (2015). Stereotype threat, mental arithmetic, and the mere effort account. *Journal of Experimental Social Psychology, 61*, 19-30.

Seitchik, A.E., Jamieson, J.P., & Harkins, S.G. (2014). Reading between the lines: Subtle stereotype threat cues can motivate performance. *Social Influence, 9*, 52-68.

Shapiro, J.R., Williams, A.M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology, 104(*2), 277-288.

Shewach, O.R., Sackett, P.R., & Quint, S. (2018, April). *The effect of stereotype threat in operational test settings: A meta-analysis.* Paper presented at the 33rd annual conference of the Society for Industrial & Organizational Psychology, Chicago, IL.

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*(1), 99.

Smith, J.L., & White, P.H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*(6), 1040-1057.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Steel, P., & König, C.J. (2006). Integrating theories of motivation. *The Academy of Management Review, 31*, 889-913.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613.

Steele, C.M. (2010). *Whistling Vivaldi: And other clues to how stereotypes affect us*. New York: Norton.

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.

Steele, C. M., & Davies, P. G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16,* 311–326.

Steele, C.M., Spencer, S.J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.

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

Stone, J., Perry, W., & Darley, J.M. (1997). "White men can't jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology, 19*(3), 291-306.

Taylor, L.A. (1987). Decision quality and commitment within a probabilistic environment. *Organizational Behavior and Human Decision Processes, 39*(2), 203-227.

Taylor, L.A., Cosier, R.A., & Ganster, D.C. (1992). The positive effects of easy goals on decision quality and risk propensity in an MCPLP task. *Decision Sciences, 23*(4), 880-898.

Thompson, E.R. (2007). Development and validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227-242.

Tolli, A.P., & Schmidt, A.M. (2008). The role of feedback, causal attributions, and self-efficacy in goal revision. *Journal of Applied Psychology, 93*(3), 692-701.

Trougakos, J.P., Beal, D.J., Cheng, B.H., Hideg, I., & Zweig, D. (2015). Too drained to help: A resource depletion perspective on daily interpersonal citizenship behaviors. *Journal of Applied Psychology, 100,* 227-236.

Unsworth, K., Yeo, G., & Beck, J. (2014). Multiple goals: A review and derivation of general principles. *Journal of Organizational Behavior, 35,* 1064-1078.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635-654.

Vancouver, J.B., More, K.M., & Yoder, R.J. (2008). Self-efficacy and resource allocation: Support for a nonmonotonic, discontinuous model. *Journal of Applied Psychology, 93,* 35-47.

Vancouver, J.B., & Purl, J.D. (2017). A computational model of self-efficacy's various effects on performance: Moving the debate forward. *Journal of Applied Psychology, 102*(4), 599-616

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*(3), 506.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*(4), 605.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement, 57*, 995-1015.

von Hippel, C., Issa, M., Ma, R., & Stokes, A. (2011). Stereotype threat: Antecedents and consequences for working women. *European Journal of Social Psychology, 41*(2), 151-161.

Vroom, V.H. (1964). *Work and motivation.* New York: Wiley.

Walton, G.M., Murphy, M.C., & Ryan, A.M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 523-550.

Watson, D., Clark, L., & Tellegen, 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.

Weidman, A.C., Steckler, C.M., & Tracy, J.L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion, 17*(2), 267-295.

Wood, R.E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37*, 60-82.

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. W. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. *Journal of Personality and Social Psychology, 103*(4), 635-646.

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*, 12–19.

Yeo, G.B., Frederiks, E.R., Kiewitz, C., & Neal, A. (2014). A dynamic, self-regulatory model of affect and performance: Interactions between states, traits and task demands. *Motivation and Emotion, 38*(3), 429-443.

Yeo, G.B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*(2), 231-247.

Yeo, G.B., & Neal, A. (2008). Subjective cognitive effort: A model of states, traits, and time. *Journal of Applied Psychology, 93*(3), 617-631.\

Zigerell, L.J. (2017). Potential publication bias in the stereotype threat literature: Comment on Nguyen and Ryan (2008*). Journal of Applied Psychology, 102*(8), 1159-1168.