“Am I Math Compatible?”: How Stereotype Threat Relates to Gender and Math Identification in Women

Elyse Murphy Neubauer
Bard College

"Am I Math Compatible?":

How Stereotype Threat Relates to

Gender and Math Identification in Women

Senior Project Submitted to

The Division of Social Studies

of Bard College

by

Elyse Neubauer

Annandale-on-Hudson, New York

April, 2014

For my parents, Karyn and Karl. Thank you for getting upset when I came home from elementary school announcing, “I don’t have to be good at math because I have an imagination”. Your annoyance directed at this comment was extremely helpful.

Acknowledgements

I would like to thank my adviser, Kristin Lane, for her hard work during this process. I would not have been able to complete this project without her essential input and encouragement. Thanks also to Barbara Luka and Sarah Dunphy-Lelii for providing me with valuable feedback. I am also grateful to Beth Hoffman-Patalona for her skills recruiting participants. I can’t believe we got 71 participants in one month! Finally, a very special thank you goes out to the best brother of all time, Nathan Neubauer.

Table of Contents

Abstract

The current study investigates the relationship between stereotype threat (i.e. adverse effects of stereotypes on stereotyped individuals' performance in stereotype-relevant tasks), gender identification, and math identification in women completing a mathematics test. Prior to test-completion, stereotype threat was induced by telling female participants that men had previously out-performed women on the test. Unthreatened women were told that both men and women had performed equally. Participants completed measures of implicit and explicit gender and math identification and rated traits on how typical they were of "male" or "female." I hypothesized that stereotype threat would increase both gender and math identification, but that perceptions of the female gender would be different for women who were highly math identified. Stereotype threat *did* occur in the current sample, who reported fairly egalitarian attitudes about gender compared to samples in previous research. There was a non-significant trend for women's female identification to increase in the face of stereotype threat. Interestingly (and counter to predictions) women under threat reported lowered identification with math compared to unthreatened women. In addition, math identification did not seem to have an effect on how the female gender was perceived by women. Implications and limitations are discussed.

"Am I Math Compatible?":

How Stereotype Threat Relates to Gender and Math Identification in Women

GIRLS
Can build a spaceship,
GIRLS
Can code a new app,
GIRLS
Should grow up knowing
That they can engineer THAT.

-Goldie Blox ™ Commercial (original)[1]

Girls should grow up without a fear of math and technology. This is the main message of www.goldieblox.com, a company providing toys designed to increase girls' interest in engineering (Goldie Blox ™, 2014). The commercial features a group of girls, assumed to be bored by the same old toys (i.e. pink princess dresses, wands and tiaras) who build a complex chain reaction out of old toys, chairs and other household objects. The existence of the commercial highlights the perception that girls must be encouraged to be excited about mathematics and logical reasoning because they do not believe themselves to be naturally good at it.

Research indicates that this type of ad-campaign is, indeed, extremely necessary. Women are generally underrepresented in Science, Technology, Engineering, and Mathematics (STEM) fields. Perhaps unsurprisingly, the gap is most obvious in mathematics. The current study was designed to understand more about this gender gap

[1] The commercial currently featured on www.Goldieblox.com no longer has lyrics explaining their position. However, the original version can be found on YouTube at, https://www.youtube.com/watch?v=rvUguiKlELg, as of 4/20/14.

and how the stereotype that women are bad at math perpetuates this divide. Although similar gender gaps are most likely prevalent around the world, the research here is based on the gender stereotype in the United States. The influence of the gender stereotype on women's personal identity was also explored.

At a young age, there are no discernible gender differences on standardized tests of mathematical ability. Average math scores of 4th grade girls and boys on the National Assessment of Educational Progress (NAEP) test never differed by more than three points at all nine testing periods occurring between 1990 and 2011 (National Science Foundation, 2014a). Similarly, a gender difference was not apparent in the NAEP scores of 8th grade students. These scores did not differ by more than two points across that same time span (National Science Foundation, 2014a). Clearly, a gender difference in math performance does not exist in middle and elementary school. In other words, girls do not seem to have trouble understanding math relative to boys.

However, as children grow into young adults, less and less women are represented in mathematics and math-related fields. In 2012, male freshmen intent on majoring in mathematics, statistics or computer science made up about five times the amount of women with the same intentions (women=1.4%, men=5.2%) (National Science Foundation, 2014b). In addition, 18.3% of intended engineering majors were also male while 3.9% were female (National Science Foundation, 2014b). Although women edged out men by a fair amount in intended biological/agricultural science majors (women=14.3%, men=10.8%) and social/behavioral sciences (women=12%, men 8.4%), intentions to major in physical science showed a male bias (women=1.9%, men=3.1%) (National Science

Foundation, 2014b). This is not altogether surprising as physical sciences, such as physics and chemistry, require more intense study in mathematics from the outset than biological or social sciences. This gender gap also persists as education increases. In 2011, 57.02% of mathematics/statistics majors who earned bachelor's degrees were men (National Science Foundation, 2014c). In that same year, a greater majority of men (64.88%) went on to enroll in graduate school (National Science Foundation, 2014d).

Why is this? What happens between elementary school and graduate school that creates this gender gap in mathematics? A recent article in the *New York Times* provides a personal account of what it is like to be a woman studying a mathematics-based subject (i.e. physics). The author, Eileen Pollack, remembered feeling unwelcome in her undergraduate physics classes. Although she was at the top of her class senior year, she did not apply to graduate school because she was not encouraged to do so (Pollack, 2013). Her personal experience is punctuated by the results of a recent study on faculty member bias. Science faculty members at six American colleges and universities showed a clear gender bias in favor of men when evaluating applications for a research assistant position. Applications with the name John at the top were evaluated more positively (applicants were even offered more money) than when the name was changed to Jennifer (Moss-Racusin, Dovidio, Brescoll, Graham & Handelsman, 2012). The *New York Times* article went on to discuss how women are viewed negatively in mathematics and science. These negative perceptions seem to influence not only outside perception, but also the perception of the self. Pollack writes "I was tired of dressing one way to be taken seriously as a [physical] scientist while dressing another to feel feminine." Understanding more about this struggle with apparently conflicting personal identities may provide a clearer picture of how stereotypes affect women's behavior.

In the case of women in mathematics, the stereotype that men are innately better at math can not only affect the way women are treated and their self-perception, but can also alter their behavior. Critically, the gender stereotype influences women's mathematical performance. Making the gender stereotype salient decreases a woman's ability to perform as successfully as possible on a math test. This is a phenomenon known as stereotype threat. Stereotype threat works primarily through increasing levels of anxiety. Imagine being asked to learn a dance in a room full of people with signs on their chests that say "experts", while you wear one that reads "beginner". Anxiety about each wrong step would be much stronger than if this distinction was not made or if it could easily be forgotten. Stereotype threat research shows that this anxiety actually creates a performance deficit, understandably affecting an individual's ability to carry out the task at hand. In the case of the typical female undergraduate studying math, anxiety about the gender stereotype causes a decrease in her ability to work most effectively.

Stereotype threat effects have been established in previous research, but the relationship between female gender identification and stereotype threat has not been fully flushed out. Since the stereotype about women in math involves gender, it follows that this type of stereotype threat and gender identification should be linked. Gender is an example of a label (such as the "expert" or "beginner" dichotomy) that is often an essential part of personal identity. On average, people had strong identification with their self-designated gender (Lane, Goh & Driver-Linn, 2011; Nosek, Banaji & Greenwald, 2002). This strong identification happens on both less-conscious (Lane et al., 2011, Nosek et al. 2002, Nosek & Smyth, 2011) and more-conscious (Lane et al., 2011; Nosek et al., 2002) levels. “Female”, therefore, is both an essential part of the gender stereotype and personal identity for women. However, as was illustrated by Eilleen Pollack’s comment about dressing to feel more like a scientist than feminine, gender is not the only type of personal identity at work during stereotype threat. In fact, the gender stereotype creates clashing identities between gender and math related perceptions of the self. With gender identity already so strong, I hypothesize that when women’s math identities are threatened by the stereotype, their gender identities will be strengthened. I also hypothesize that this can occur even for women who highly identify with math, although their ideas of what “female” means may be different from women who do not identify as highly with math.

Stereotypes and Stereotype Threat

Before tackling the relationship between stereotype threat and gender identification, stereotype threat will be more thoroughly explained. At the heart of stereotype threat is, of course, the stereotype. In this case, the stereotype is that men are better at math than women. It is not imperative for women to explicitly agree with this statement in order for stereotype threat to work. However, it is necessary for some amount of stereotype internalization to occur, which most likely happens on a less conscious level.

The Stereotype: Math=Male. In order to understand the stereotype threat effect for women in mathematics, the prevalence of the gender stereotype must first be explored. Both men and women hold the stereotypes that men are better at math than women and that women are better at languages and arts than men (Kiefer & Sekaquaptewa, 2007;

Nosek et al., 2002). Furthermore, this internalization of the gender stereotype has been measured on both explicit (Kiefer & Sekaquaptewa, 2007) and implicit (Nosek et al., 2002; Nosek & Smyth, 2011) levels. An explicit attitude is an attitude or belief that is held consciously and can be measured through self-reports in which individuals are specifically asked about a belief. Implicit attitudes, on the other hand, are beliefs that are held below the surface. Individuals may be consciously suppressing, be unaware of, or simply less aware of these attitudes.

Men and women implicitly (i.e. less consciously) associated “male” with “math” on the computerized Implicit Association Test (IAT) (Nosek et al., 2002; Nosek & Smyth, 2011). The IAT is a measure of implicit attitude used in many of the studies investigating the gender stereotype in mathematics (Lane et al., 2011; Nosek et al, 2002; Nosek & Smyth, 2011). The IAT measures the strength of the association between words. In the computerized version, if individuals are faster at categorizing certain words together (i.e. “male” words into the “math” category) than it is concluded that they have a greater association between those concepts (Greenwald, McGhee & Schwartz, 1998). Participants who were quicker at categorizing “male” with “math” than “female” with “math” internalized the gender stereotype at the implicit level (Nosek et al., 2002; Nosek & Smyth, 2011).

The stereotype can also be measured on an explicit (i.e. conscious) level, although this type of measurement yields a greater variation of results. A sample of 138 females reported the percentage of men and women who they believed fit certain characteristics. Women reported greater percentages of men fitting the “good at math” trait, providing evidence for a generally strong endorsement of the gender stereotype in math on a more explicit level than the IAT (Kiefer & Sekaquaptewa, 2007). However, when asked even more explicitly to indicate level of agreement with statements such as "Men are better at math than women are.", both men and women reported low levels of agreement (Nosek & Smyth, 2011). Interestingly, men reported higher levels than women. It is possible that people are less likely to express endorsement of the stereotype if blatantly asked about this endorsement.

In order to understand why more variation of stereotype endorsement exists when measuring the stereotype explicitly, it is helpful to understand the predictive validity of implicit and explicit measures. In a meta-analysis which reviewed the predictive validity of the IAT and self-report measures in 122 published articles, it was concluded that the use of implicit and explicit measures had independent benefits (Greenwald, Poehlman, Uhlmann & Banaji, 2009). The IAT, in particular was the best at predicting interracial and intergroup behavior. However, in less socially sensitive domains (i.e. consumer and political preferences) self-report measures had better predictive validity (Greenwald et al., 2009).

Difference in academic ability based on gender is a socially sensitive topic. It is, therefore, not surprising that the gender stereotype can be consistently found using implicit measures. Results using explicit measures are less consistent. Indeed, men and women who were tested on their implicit and explicit endorsement of the gender stereotype did not explicitly report that they thought women were worse at science than men. However, both men and women showed stronger implicit associations between male and science than they did between women and science (Lane et al., 2011). Implicit attitudes have been shown to be superior predictors of women's plans to pursue science (Lane et al., 2011) and both genders' level of engagement in math (Nosek & Smyth, 2011). Critically, implicit measures were better at predicting this engagement for women (Nosek & Smyth, 2011).

Men and women's attitudes towards math also reflect this stereotype. Men and women's implicit attitudes towards math were measured and compared to implicit attitudes towards language and arts. In each comparison, men and women had negative attitudes towards math, but women's attitudes were *more* negative. (Nosek et al., 2002; Nosek & Smyth, 2011). This attitude pattern was reflected in implicit self-evaluations. Women matched words representative of the self with "art" faster than with "math". This suggests that women's concept of themselves was more strongly identified with art than math. Men did not identify more strongly with either subject. Interestingly, a gender division was not detected when measuring the implicit acceptance of the gender stereotype in math. More simply put, both men and women showed implicit internalization of the stereotype (Nosek et al., 2002).

Even though men and women do not always explicitly endorse the gender stereotype, explicit measures sometimes mirror the pattern found by implicit measures (Nosek & Smyth (2011). On an explicit level, the same women who did not report endorsing the stereotype reported less math identity, higher math anxiety, less math participation, less math ability and less warmth towards math than men (Nosek & Smyth (2011)). This creates more support for the idea that the stereotype is internalized, even if participants are not always consciously aware of the fact or comfortable reporting endorsement of the stereotype. This not only shows that the gender stereotype is held by the average American man or woman, but that women perceive their gender as less connected to the math world, while men do not. Internalizing the gender stereotype is related to more serious deficits in math-connectedness for women than for men.

Stereotype Threat. One of the main mechanisms behind the gender gap in math is stereotype threat. Again, this threat refers to the negative influence stereotypes have over performance (Barber & Mather, 2013; Ben-Zeev, Fein & Izlicht, 2005; Ben-Zeev, Dennehy, Sackman, Olide & Berger, 2011; Cohen & Garcia, 2005; Kiefer & Sekaquaptewa, 2007; Lesko & Corpus, 2006; Martens et al., 2005; ; Pronin, Steele & Ross, 2004; Spencer, Steele & Quinn, 1999; Steele & Aronson, 1995; Steele, 1997; Wout, Danso, Jackson & Spencer, 2008; Steinberg, Oken & Aiken, 2012; Stone, Lynch, Sjomeling & Darley, 1999 ; Zhang, Schmader & Hall, 2012). Stereotype threat can be more specifically defined as the phenomenon that occurs when stereotype salience increases the fear of confirming that stereotype. This causes a member of the stereotyped group to perform worse on a task perceived to measure the stereotype. More specifically, if a woman was told she was about to take a math test on which women traditionally do poorly or that she was about to take a test able to accurately measure her math ability, she was more likely to do badly on that test than if she was not given such cues (Ben-Zeev et al., 2005; Ben-Zeev et al., 2011; Cohen & Garcia, 2005; Kiefer & Sekaquaptewa, 2007; Lesko & Corpus, 2006; Martens et al., 2005; Pronin et al. ,2004; Spencer et al., 1999; Steinberg et al., 2012 ; Wout et al., 2008; Zhang et al., 2012). A subtle reminder of the gender stereotype was enough to hinder math performance.

Stereotype threat is a well-documented phenomenon that does not only exist for women in mathematics. African Americans who were told that verbal GRE questions were designed to test their academic abilities received lower scores than African Americans who were told that the questions were unrelated to academic ability (Steele & Aronson, 1995) Furthermore, Black participants who were under the impression that the test was unrelated to academic ability performed equally as well as White participants (Steele & Aronson, 1995). Stereotype threat can occur for elderly adults as well. When elderly adults read an article about memory decline in old age and were then asked to report their age, they performed worse on a memory task than if they read about increased memory capabilities in old age (Barber & Mather, 2013). Stereotype threat has also been documented in White men, resulting from the stereotype that White men do not have the same innate athletic ability as Black men (Stone et al, 1999).

As was mentioned above, one of the factors related to the strength of stereotype threat is an internalization of the stereotype itself. Internalization of the stereotype that men are good at math while women are not is, unsurprisingly, linked to women's susceptibility to stereotype threat within math and science. Stereotype threat in mathematics can, therefore, be explained a little differently. Women who were compelled to think about the gender stereotype tended to show lower scores than women who were not compelled to think about the stereotype (Ben-Zeev et al., 2005; Ben-Zeev et al., 2011; Cohen & Garcia, 2005; Kiefer & Sekaquaptewa, 2007; Lesko & Corpus, 2006; Martens et al., 2005; Pronin et al., 2004; Spencer et al., 1999; Steinberg et al., 2012; Wout et al., 2008; Zhang et al., 2012).

Stereotype threat works through anxiety. Specifically, higher arousal levels have an effect on task performance. At higher arousal levels, difficult tasks become more difficult, while easy tasks become less challenging. Women under stereotype threat performed worse on a difficult task, but better on an easy task than their non-threatened counterparts (Ben-Zeev et al., 2005). This means that women's poor performance under threat was due to higher levels of arousal or agitation caused by being reminded of the stereotype.

In order for anxiety to be produced in this way, an individual must believe in the stereotype (either implicitly or explicitly). Otherwise, thinking about the stereotype would not increase arousal levels and would, in turn, have no effect on behavior. Indeed, women who reported higher levels of stereotype internalization had higher performance under the non-threat condition, than women who reported lower levels of stereotype internalization (Kiefer & Sekaquaptewa, 2007). Interestingly, while the initial level of stereotype internalization is important, being put under threat strengthens the effects in women who did not previously endorse it. Women who were most likely to prove the stereotype wrong (i.e. women low in stereotype internalization) were brought down to the performance level of women with high stereotype internalization when put under threat (Kiefer & Sekaquaptewa, 2007). This suggests that women who are already high in stereotype internalization are constantly under some sort of threat, while other women's internalization is strengthened by being put in threat situations.

If stereotype threat causes poor performance through increased anxiety (Ben-Zeev et al., 2005), then it is an increase in anxiety that brings women under threat with low stereotype internalization to the performance level of those with high stereotype internalization. Theoretically, this means that in every situation in which a woman who has remained relatively stereotype-free is reminded of the gender stereotype, this woman's anxiety is increased. Feeling especially anxious while taking a test, not to mention performing badly on it compared to men, must provide perceived "proof" of the stereotype. This, in turn, would increase stereotype internalization, creating more fuel for anxiety until the woman has internalized the stereotype enough to be constantly under some sort of threat.

At this point it is clear that women have more negative attitudes towards math than men (Nosek et al., 2002), that women are less likely than men to identify themselves with math (Nosek et al., 2002) and that women are negatively affected by stereotype threat during math tests (Ben-Zeev et al., 2005; Ben-Zeev et al., 2011; Cohen & Garcia, 2005; Kiefer & Sekaquaptewa, 2007; Lesko & Corpus, 2006; Martens et al., 2005; Pronin et al., 2004; Spencer et al., 1999; Steinberg et al., 2012; Wout et al., 2008; Zhang et al., 2012) . It is also clear that, while initial internalization of the stereotype affects the strength of stereotype threat effects, being under threat also strengthens the original stereotype (Kiefer & Sekaquaptewa, 2007). This suggests a cycle in which stereotype threat increases the likelihood of poor math performance in women. Remember, the typical female undergraduate has internalized the gender stereotype in some way (Nosek et al., 2002). If she is low in stereotype internalization, it is possible that being in math classes creates a stereotype threat situation in which her internalization of the stereotype is strengthened every day. If she is high in stereotype internalization, she may be under some sort of math related threat at all times, but especially in math settings. Regardless of stereotype internalization level, increased exposure to mathematics leads to increased stereotype threat effect. This suggests that as long as the gender stereotype exists, all women are susceptible to threat effects.

Identity

Besides internalization of the stereotype itself, women's vulnerability to stereotype threat is influenced by two remaining factors. These factors are identification with gender and identification with math. In sum, as gender and math identification increase, the effects of stereotype threat tend to get stronger and math performance, poorer. However, women who are higher in gender identification tend to be lower in math identification and vice versa. This is interesting because these two types of identification seem to have the same relationship with stereotype threat strength, but opposite relationships with each other. In order to understand this confusing pattern of results, it may be helpful to provide a brief understanding of personal identity.

Personal Identification/ Group Identification. According to social identity theory, personal identity is built up of multiple group affiliations that are inevitably compared to other groups throughout a lifetime (Hornsey, 2008). Some of the first experiments regarding group identification assigned participants arbitrary groups with no social context (i.e. by the flip of a coin). Attachment to the assigned group was so strong that participants, who had no contact with out-group members, allocated a greater number of points to the in-group even though this provided no personal benefit (Turner, 1978 , as cited in Hornsey, 2008). Social identity theory also states that the group to which an individual belongs is treated more favorably in an effort to boost the individual's self-image (Hornsey, 2008), suggesting that personal identity is linked to group membership.

Considering social identity theory in the context of stereotype threat, one might think that a member of a stereotyped group would not possess high self-esteem. Since their group membership (i.e. their personal identity) is being threatened, self-esteem might decrease. However, stereotyped individuals' self-esteem is usually the same as their non-stereotyped counterparts (see Crocker & Major, 1989). In one example, black students who were put under stereotype threat did not report different levels of personal worth regarding academics (i.e. "I feel as smart as others") than Black students not under threat and White students (Steele & Aronson [study 1], 1995). One theory as to why this is posits that members of stigmatized groups are selective in what group characteristics are connected to their self-esteem. Continuing with this line of thinking, a stigmatized individual would be less likely to see negative stereotypes as connected to their personal worth (see Crocker & Major, 1989). Women's negative evaluations of math and their decreased likelihood of identifying with math compared to men are examples of this disconnection. Perhaps women are not connected to math because rejecting mathematics as an important factor in their lives can serve as a form of self-preservation. Theoretically, if math is not included as a part of personal identification, stereotype threat in math should not affect satisfaction with the self. Furthermore, if self-esteem is protected in such a way, it would be most advantageous for personal identification to be changeable. If this is the case, identification may subtly change according to which identification would most likely protect self-esteem in any given situation.

Such a theory would suggest that women under stereotype threat would be less identified with math. Unfortunately, research does not paint such a simple picture. As is illustrated by Figure 1, the strength of women's gender and math identities have similar relationship with stereotype threat. Although clear causal relationships have not yet been established, increased gender and math identity are both related to increases stereotype threat. However, math identification and gender identification are inversely related. Before we grapple with this continuity problem, gender identification and math identification will be discussed separately.

Are Gender Identification and Stereotype Threat Really Related?. Previous research suggests that women who strongly identify with their own gender are more susceptible to stereotype threat (Ben-Zeev et al., 2011; Cohen & Garcia, 2005; Lane et al., 2011; Nosek et al., 2002 Wout et al., 2008). When women identified more with their own gender, their identification with math decreased and they had more negative attitudes towards math than women who were not as strongly identified with being female (Nosek et al., 2002). Similar findings exist in relation to gender stereotypes in science. The gender stereotype in science was more pronounced for women who strongly identified with female (Lane et al., 2011). A clear pattern is forming, in which women who see themselves as feminine do not see themselves as science or math oriented. "If math (or science) equals male and male does not equal me, than I do not equal math (or science)" (Nosek et al., 2002). It follows that women who are not connected with math, and therefore more connected with female, may be more likely to internalize the stereotype that math is simply for men and not for women. Since internalization of the stereotype is related to the strength of stereotype threat effect, gender identification should be similarly related to stereotype threat strength.

In accordance with this line of thinking, women with higher gender identification performed worse under stereotype threat than women who were low in gender identification (Wout et al. [study 2], 2008). In this study, the type of stereotype threat manipulation was controlled for. Women were either under group-threat, which specifically evaluated the math performance of their gender, or self-threat, which specifically evaluated their own math performance unconnected to group membership. Only women whose gender was highlighted (i.e. they were told that all scores would be compared by gender and individual tests were put into a pile of tests marked "female") showed an increase in gender identification as stereotype threat effects increased. When women were threatened without highlighting gender (i.e. they were told that they would grade their own test and that no one else would know their score) the effects of stereotype threat did not increase with gender identification (Wout et al. [study 2], 2008). This is very important, as it shows that the effect on performance is enhanced due to anxiety about harming the group's (i.e. women's) reputation. When gender is highlighted, identifying with being a woman means identifying with the stereotyped group. Therefore, if a woman's score is being evaluated in the context of gender, the manipulation should increase her anxiety and decrease her performance relative to this identification.

This provides direct evidence that stereotype threat strength depends on gender identification. However, because the type of stereotype manipulation was controlled for, it is unclear if more generic types of stereotype threat manipulations would yield the same effect. It may be that the stereotype threat effects found in other studies are caused largely by group-threat (i.e. manipulations that mention gendered groups). The manipulation highlighting gender was not completely isolated in other studies as it was in Wout et. al.(2008). Fortunately, gendered groups were specifically mentioned in most manipulations allowing the results of these manipulations to be compared to Wout et al.'s (2008) findings. Such manipulations include telling participants that stereotype-congruent gender differences were found in previous studies (Lesko & Corpus, 2006; Martens et al. [study 2], 2006; Pronin et al. [studies 2 & 3], 2004; Spencer et al., 1999; Steinberg et al., 2012; Zhang et al., 2012) and putting highly math identified women into groups in which they were the only female (Ben-Zeev et al., 2005).

This is not to say that stereotype threat does not work when gendered groups are *not* mentioned. Women's math scores can be lowered by a stereotype threat manipulation that does not specifically mention gender. This manipulation is usually done by making it clear that the test is designed to be an accurate test of math ability. When this method, and variations of it, are used women still perform worse on a math test when under stereotype threat (Wout et al. [control conditions], 2008; Martens et al. [study 1], 2006; Cohen & Garcia, 2005). While these manipulations produce stereotype threat effects, they may not be related to changes in gender identification because they do not mention gendered groups. If this is the case, than Wout et al.'s (2008) findings about gender identification should be replicated by using any stereotype threat manipulation which mentions gender. In other words, if the manipulation mentions gender, an increase in gender identification should be related to an increase of stereotype threat effects.

Even with evidence of a direct relationship between gender identification and stereotype threat, discrepancies in the data between Wout et al.'s (2008) studies make definitive conclusions more difficult. There were some contradictory results found in studies 1 and 2. In study 1, women in the gendered-threat condition did not exhibit stereotype threat effects while the women in the diagnostic condition did. Besides conditions that either mentioned gender or did not, diagnostic and non-diagnostic conditions were used as controls. The diagnostic condition was the alternative stereotype threat manipulation described above in which gender differences were not mentioned, but the evaluative ability of the math test was. Again, the diagnostic condition yielded stereotype threat effects while the threat highlighting gender did not. In study 2, however women in the gendered-threat condition did exhibit stereotype threat effects. One explanation for this difference is that the person conducting experiments in study 2 was male (in study 1, this person was female). Having a male experimenter is thought to have bolstered the gendered-threat effect, lending credence to the idea that stereotype threat strength and gender identification are related.

Another explanation is that women's gender identification was measured in study 2 but not in study 1. This means that there is no way of knowing if women in study 1 who were high in gender identification were actually affected by stereotype threat. This would make sense, considering study 2 showed stereotype threat effects were stronger for women who were high in gender identification (Wout et al., 2008). It is clear that more research is needed to understand this relationship. Again, if the relationship between gender identification and stereotype threat strength depends on the use of a threat manipulation highlighting gender, then study 2’s findings should be replicated by using such manipulations.

Although the possibility that gender identification may not be related to stereotype threat must be entertained because of inconsistencies in these studies, patterns found in the literature suggest this relationship must exist. The perceived incompatibility between woman and math is clear. Considering the evidence that women tend to internalize the stereotype that men are better at math than women (Nosek et al., 2002; Nosek & Smyth, 2011) and that women's endorsement of similar stereotypes in science were moderated by gender identity (Lane et al., 2011), it seems foolish to ignore a relationship between gender identification and stereotype threat. Furthermore, a study that did not concluded gender identification was related to stereotype threat strength (Kiefer & Sekaquaptewa, 2007) used methods suggesting that the use of a stereotype manipulation highlighting gender is, indeed, critical to finding this relationship. Women in this study were given a stereotype threat manipulation similar to the diagnostic manipulation in Wout et al.'s (2008) study. Kiefer and Sekaquaptewa's (2007) manipulation did not mention gender groups.[2] Considering the influence differing stereotype threat manipulations have had over results, the relationship between gender identification and stereotype threat strength cannot be dismissed based on this research.

One of the goals of the current study is to provide more conclusive evidence of the relationship between gender identification and stereotype threat. Taking the disconnect between women and math into account, I hypothesize that these concepts are related and that Wout et al.'s (2008) findings concerning gender identification will be replicated using

[2] Kiefer and Sekaquaptewa (2007) did not provide detailed accounts of their stereotype threat manipulation. Instead the reported, "Following the procedure used by Quinn and Spencer (2001), directions described the test as either diagnostic (threat condition) or non-diagnostic (reduced threat condition) of mathematical ability." The diagnostic condition used in Quinn and Spencer (2001) was not explicit in claiming that the test was designed to accurately test math ability. However, the diagnostic condition did not mention gendered groups.

a stereotype manipulation that focuses on gender differences. More specifically, I hypothesize that women under stereotype threat will experience an increase in gender identification.

Stereotype Threat: A Causal Mechanism. For the moment, it will be assumed that gender identification and stereotype threat strength are related. This means that women who strongly identify with "female" are more likely to experience low math scores due to stereotype threat. In this case, the obvious solution to stereotype threat is to dis-identify with the group (i.e. female). I hypothesize, however, that female identification may actually be enhanced by being under stereotype threat, making dis-identification from the group much more challenging. Previous research has stopped short of concluding that stereotype threat causes such identity changes. Instead, gender identification is presented as a moderator, something that enhances the relationship between stereotype threat and test performance (Wout et al., 2008). This interpretation, although valid, may be too simplistic.

The idea that an increase in gender identification is related to an increase in stereotype threat effect for women can be interpreted in three ways. The first is that initial high gender identification strengthens the stereotype threat effect. This conclusion is hinted at in Wout et al. (2008) in which gender identification is described as a moderator. Causation, however, is not concluded as moderation implies an enhancement of the relationship rather than a strict "A" leads to "B", then "B" leads to "C" scenario. Unfortunately a clear direction of causality between gender identification and stereotype threat strength cannot be concluded from Wout et al. (2008). Although stereotype threat was induced and gender identification was measured, the order of these events was not reported. In order to conclude anything about stereotype threat as a causal mechanism, gender identification must be measured after creating stereotype threat. However, since gender identification is a moderator, it is likely that existing levels of identification affected stereotype threat.

The second interpretation is that the process of being under stereotype threat increases gender identification in women. A woman who is put under stereotype threat in mathematics becomes, at least temporarily, slightly more attached to her female identity and less attached to her mathematics identity. This is the hypothesis of the current study. This hypothesis is consistent with the theory that self-esteem of stereotyped group members is protected by increasing a personal identity that is not threatened (Crocker & Major, 1989; Steele & Aronson [study 1], 1995). In this case, a woman under stereotype threat may identify more with being a woman than being good at math because being a woman is a safer personal identity in which to place her self-value.

The third interpretation is that neither stereotype threat nor gender identification begin the causal relationship. Instead, a unidirectional relationship exists in which both types of causal relationships work to affect women's math performance depending on the woman and the situation. Given that the relationship between gender identification and stereotype threat strength itself may be easily erased through the use of different threat manipulations, the interpretation that takes changing personal and situational factors into account is most likely to be true. It is beyond the scope of this study to understand if this third interpretation is most relevant or how it may work. To date, however, there is little to no research directly claiming that the second type of relationship exists. This is the relationship in which stereotype threat causes an increase in women's gender identification. Therefore, another goal of this study is to find out if this type of causal relationship exists. Again, I hypothesize that it does, though I do not maintain that it is the only type of relationship at work in stereotype threat situations.

There are some interesting findings providing the groundwork on which this hypothesis was built. Think of the actual experience of being under stereotype threat. At the point that stereotype threat has taken effect, the stereotype itself is either already strong or being strengthened depending on a woman's initial level of stereotype internalization (Kiefer & Sekaquaptewa, 2007). At this point, women also showed signs of a possible increase in gender identification. Women under stereotype threat who were filmed during an interview were later coded on their use of flirtatious behavior (i.e. feminization). Women under threat were perceived as using more nonverbal flirtatious behavior (i.e. adjustment and touching of clothing and accessories, leaning forward, crossing legs, etc…) than women who were not under threat. This behavior may suggest that identification with feminine characteristics became hyper salient (at least temporarily) while women were under threat, possibly leading to an increase in feminized behavior and certainly causing observers to perceive feminized behavior.

There could, of course, be other explanations for an apparent increase in flirtatious behavior. Women may have been exhibiting symptoms of anxiety that were perceived as flirtatiousness. Although women's actual anxiety levels were not measured, outside observers coded for anxious behavior. Coder perceived anxiety and coder perceived flirtatiousness did not positively relate to each other (Ben-Zeev et al., 2011). This suggests that although stereotype threat is reported to work through increasing anxiety (Ben-Zeev et al., 2005), this anxiety had not played a direct role in increasing flirtatious behavior. If anxiety and flirtatiousness had been coded in a way that was less influenced by the subjectivity of outside observers, these findings would provide more concrete evidence for the hypothesis that stereotype threat caused an increase in feminized behavior. As it is, there are many alternative explanations for these results. Women's could be exhibiting anxious behavior. In addition, all people under threat may simply appear less confident and more submissive, regardless of gender. The possibility still remains, however, that increased flirtatious behavior is linked to increased femininity and that this occurs through stereotype threat. The results of our current study will shed more light on whether or not this interpretation is valid.

Strengthened gender identification was also observed in women who simply witnessed another woman undergo stereotype threat. In this case, the woman who witnessed stereotype threat were more likely to characterize themselves as stereotypically female using Pronin et al.'s (2004) scale of female stereotypic character traits (i.e. "gossipy", " leaving work to raise children") (Cohen & Garcia [study 3], 2005). Women who arrived for a study were told to wait while the participant before them finished. While the real participant thought they were waiting for their experiment to start, the fake participant (a female confederate) was given a stereotype threat manipulation. In the threat condition, the female confederate expressed how bad she was at math and was assured that she would be evaluated on her performance. Interestingly, the threat manipulation did not mention gender differences. However, seeing a woman distressed about math seemed to have been enough to trigger associations about gendered groups and cause stronger female characterization. If simply witnessing a woman in a stereotype threat scenario can elicit stronger feminine identification, it stands to reason that the effects are even more robust when a woman is under stereotype threat herself.

Math Identification. Previous research also shows a relationship between high levels of math identification in women and stronger stereotype threat effects (Steinberg et al., 2012). The most plausible theoretical explanation for this is that women who identify with math should value excellent math performance much more than women who do not identify with math (Spencer et al., 1999; Steele., 1997). Therefore, failure should have more serious consequences for women who highly identify with math then for women who do not. Considering that stereotype threat works through increasing levels of arousal in women (Ben-Zeev et al., 2005), it follows that anxiety levels brought on by stereotype threat may be intensified in women who have more to lose by failing.

There is relatively little research done on the different effects of stereotype threat due to varied levels of math identification. The research that does exist tells us women who were under stereotype threat showed lower math scores if they highly identified with math and had high GPAs in calculus classes. Women who did not identify with math and had low calculus GPAs had higher performance levels (Steinberg et al., 2012). Women who completed pre and post-test measurements of math identification, described as math self-esteem, were also more likely to exhibit discounting behavior if they highly identified with math (Lesko & Corpus, 2006). This means that they were more likely to report that the test did not accurately measure their abilities. Perhaps women who highly identified with math exhibited more discounting behavior because they felt more anxiety during stereotype threat, leading them to provide reasons for their poor math performance. Women's math identification did not seem to decrease after Lesko and Corpus' (2006) stereotype manipulation, which does not contradict the findings of Steinberg et al. (2012). Unfortunately, the discounting study did not report what kind of relationship existed between math identification and stereotype threat strength (Lesko & Corpus, 2006). However, the available evidence suggests that an increase in math identification is related to an increase in stereotype threat strength.

As with gender identification, the directionality of this relationship is challenging to pinpoint. It is impossible to tell whether women's math identification is increased by stereotype threat in Steinberg et al. (2012). Considering the measures of math identification were taken before the stereotype threat manipulation, the only conclusion that can be made is that women who were already highly identified with math were more susceptible to stereotype threat. That being said, if stereotype threat can cause an increase in gender identification and both increased amounts of gender identification and math identification are related to an increase in stereotype threat strength, it follows that math identification should be increased by stereotype threat as well. The main goal of this study, however, is to understand the relationship between gender identification and stereotype threat. While math identification is an extremely important factor, the causal relationship between it and stereotype threat strength is not the main focus of this study.

Math Identification v.s. Gender Identification. Math identification and gender identification are inversely related in women. In other words the idea of a woman being extremely "mathy" seems to be incompatible with being extremely feminine. As noted above, women who more highly identified with being a woman tended to hold stronger gender stereotypes around math and science (Lane et al., 2011; Nosek et al., 2002). Women who were highly identified with their gender also show more performance deficits under stereotype threat than woman less identified with their gender when gendered groups were made salient in a stereotype threat manipulation (Wout et al., 2008). Women also implicitly identified with arts and language as opposed to math and had more negative attitudes towards math than they did towards arts and language (Nosek et al., 2002). Since men and women both hold the stereotype that arts and language are more associated with women than men (Nosek et al., 2002), it stands to reason that women are identifying with a domain (i.e. arts/languages) that more closely relates to female. In other words, the more feminine you are the worse you must be at math. The take home message provided by this stereotype is that "math" and "female" are perceived as incompatible.

Supporting this observation, women who identified more with math were less likely to view themselves as stereotypically feminine. Specifically, they were less likely report that feminine character traits perceived as "math-irrelevant" describe themselves (Pronin et al., 2004). An example may help here. In an interview based study, 10 women majoring in STEM were asked about their experiences regarding science and math based college courses. Most women felt that there was a contrast between what people perceived scientists to look like and what women were stereotyped to look like. One woman remembered being told that she didn't look like a science major, and immediately assuming that this was the case because she wore makeup (Goldman, 2012). The results reported in

Pronin et al. (2004) mirror the information from these interviews. In a pilot study, some stereotypically feminine character traits were perceived as "math-irrelevant" by participants and some were not. Those that were "math-irrelevant" included traits like "wearing makeup" and "gossipy". These were the same traits rejected by highly math identified women under stereotype threat (Pronin et al., 2004).

Research also suggests that women's romantic goal pursuit is not easily merged with higher math or science identities. Women who were primed with romantic goals (i.e. looked at pictures designed to activate associations about romance or overheard a conversation about a recent date) had less interest in STEM in general, a weaker preference for STEM related majors and reported a preference for English and Foreign language (i.e. stereotypically female majors) (Park, Young, Troisi & Pinkus [studies 1, 2a & 2b], 2011). In contrast, men's interest in and preference for specific majors were not influenced when primed with romantic goals (Park et al. [studies 1, 2a & 2b], 2011). Furthermore, women who were asked to keep a diary of goals and activities reported less math related activities (i.e. working on math homework) and more romance related activities (i.e. texting someone of romantic interest) if they also reported pursuing a romantic goal (Park et al. [study 3], 2011).

While men possess romantic goals as well, being primed with such goals did not affect interest in science and mathematics for men. There is, therefore, something unique about being female that creates this conflict between math and romantic goals. It seems extremely possible that the images used to prime women with romantic goals (i.e. sunsets, beaches, candlelit dinners (Park et al. [studies 1 & 2a], 2011)) were also stereotypically female dating preferences. The results, therefore, could be interpreted in another way. The women who were primed with stereotypically female goals were less likely to indicate interest in STEM majors. This interpretation adds to the evidence that “math” and “female” are perceived as inversely related concepts.

Continuity Problem. Since math identification and gender identification are inversely related, shouldn't they also have differing effects on math performance under stereotype threat? Intuitively, the answer would be yes. Frustratingly, both higher levels of math identification and higher levels of gender identification were related to greater susceptibility to stereotype threat (refer to Figure 1). How can two factors which are inversely related to each other have the same and not opposing relationships with stereotype threat strength?

There are two possible answers to this question: 1) Previous research has used mostly explicit measures of math identification and gender identification. Perhaps more intuitive patterns will be discovered using implicit measurements. 2) The division of feminine traits in descriptions of highly math identified women (Pronin et al., 2004) suggests a division of feminine identity. In other words, one can identify as either a “math-relevant” female or “math-irrelevant” female. This would allow women high in math identification, previously shown to reject “math-irrelevant” characteristics (Pronin et al., 2004), to have a higher gender identification as a “math-relevant” female.

Simply put, women who are high in math identification would have a different understanding of what it means to be female than women who are low in math identification. This would allow both women with high math identification to have high gender identification while under stereotype threat scenarios. This suggests that when women with high math identification identify with their gender, their perception of "female" is more relevant to math than women with low math identification. If this is the case, then it may be possible to have both high gender identification and high math identification while under stereotype threat (refer to Figure 2 for a diagram of this theory).

The second main goal of this study is to test the theory that highly math identified women have different perceptions of "female" while under threat. At first, this explanation doesn't seem to take into account that highly math-identified women are still more affected by stereotype threat than women who have less identification with math. Why would women who were high in math identification still be affected by stereotype threat if they perceive their gender as "math-relevant"? In this case, math would equal "math-relevant" which would equal "female", therefore alleviating threat effects. Considering highly math identified women are not immune to stereotype threat, I propose that while characteristics describing a "math-relevant" female are not negatively stereotyped with math, they do not seamlessly fit together with the concept of math. The "math-relevant" traits used in Pronin et al. (2004) are still stereotypically female and have more to do with nurturing and caretaking than with problem solving or analytic skills. It seems that these characteristics can coincide with math but are not seen as irrefutably attached to math success. I propose that this is the case because characteristics that are perceived to fit seamlessly with math are also stereotypically male. A more accurate equation would read; math equals male, male does not equal "math-relevant" female, therefore, "math relevant" female still does not equal math. The effects of stereotype threat are maintained while the theory that women who are highly math identified may perceive their gender differently under threat remains possible.

Overview of Current Study

The current study examined how certain factors relate to the strength of stereotype threat effects. These factors are math identification and gender identification. Men and women took a math test after being exposed to either a gendered threat or non-threat manipulation. After that, their implicit and explicit gender and math identification were measured. Participants also filled out a questionnaire designed to find out how "female math-irrelevant" character traits were perceived relative to "female math-relevant" and "hypothesized-male" traits. These traits were rated on how descriptive they were of the male and female genders.

I predicted that women's gender identification would increase through the process of being under stereotype threat, but that women high in math identification would judge their gender differently than women low in math-identification. Specifically, women with high math identification would judge "math-irrelevant" traits as less typical of the female gender, than "math-relevant" traits. Women low in math identification would not make this distinction in typicality ratings. I did not have specific hypotheses about implicit measures, as the majority of the current literature has utilized explicit measurements of gender and math identification. However, it was presumed likely that explicit and implicit patterns would mirror each other. If anything, hypothesized patterns were more likely to emerge in the implicit measurements because implicit measures are more likely to pick up differences in socially sensitive subjects. My hypotheses were as follows;

Hypothesis 1: Women under stereotype threat will have higher explicit (and most likely implicit) gender identification than women not under stereotype threat.

Hypothesis 2: Under stereotype threat, women with high math identification will judge “female math-irrelevant” traits as less descriptive of their gender than “female math-relevant” traits. Women with low math identification will not differentiate between female trait types.

Method

Participants

Participants were 71 undergraduate students from Bard College, a small liberal arts college in New York State. Two participants were excluded because they failed to indicate what they thought the hypothesis of the study was in the demographic questionnaire. An additional participant was excluded because they mentioned knowledge about stereotype threat. This left a final total of 68 participants (25 Male, 39 Female). The mean age was 20.23 (SD=1.49). The sample was made up of primarily European Americans with 76.5% self-identified White/Caucasian, 4.4% African-American, 4.4% Hispanic/Latin-American, 2.9% Asian/Pacific Islander and 10.3% who specified an ethnicity not included on the questionnaire.

Recruitment primarily took place in the campus center. Since the majority of Bard students do not major in science, mathematics, or computing, it is possible that participants were not highly math identified regardless of gender. In order to insure a sample in which women who were highly math identified were represented, extra recruitment was carried out in the primary science classroom and laboratory building. Participants were recruited using a tabling technique. A table was set up in a recruiting site with a sign inviting students to participate in a psychology study in exchange for a candy bar and the chance to win $100 dollars. Participants either completed the study right away or signed up to participate on an upcoming Saturday. In order to be entered to win the $100 dollar lottery, participants provided their name and email address. Information on this lottery sheet was kept separate from the data in order to insure anonymity of data.

Materials

Stereotype Threat Manipulation. Participants were randomly assigned to either the stereotype threat condition, or the non-threat condition. Participants in the stereotype threat condition were told that they were about to take a math test that had shown differences in performance based on gender, such that men outperform women. The experimenter stated the following:

> I don't know if you are following any debates about gender and mathematics, but you may have heard the idea that men are better at math than women. Although research on this subject is ongoing, the particular math test you are about to take has revealed gender differences in previous research such that men have performed better than women on this test.

The instructions were the same for the non-threat condition except for the last sentence, which read, "Although research on this subject is ongoing, the particular math test you are about to take has *not* revealed gender differences in previous research. Men and women performed equally well on this math test."

Math Test. Participants took a math test to assess the efficacy of the stereotype threat manipulation. The math test consisted of six questions taken from GRE practice tests available in print and online (Educational Testing Service, 2007; Educational Testing Service, 2008). Stereotype threat causes higher levels of arousal which adversely affects performance on difficult tasks and improves performance on easy tasks (Ben-Zeev et al., 2005). Therefore, in order to elicit stereotype threat, the test had to be difficult enough to challenge students with varying levels of math identification and ability, but not so difficult that participants gave up. As such, it was necessary to pick questions that exhibited some variability in their difficulty level. Three questions were taken from a practice test for the quantitative sections of standard GRE practice tests and three questions were taken from a GRE mathematics subject practice test.

As part of preliminary research, volunteers (N=7) who did not participate in the current study completed questions from the standard GRE test. Three questions were chosen for the final test that showed variability in the percentage of people who got the answers right, but were not impossible to understand (percentage correct ranged from 33.33% to 77.43%). One participant in this pilot test got all the questions right with no real struggle. Considering this participant was a Physical Science major and a possible representative of a highly math identified person, it was necessary to add questions that would be harder for more math-inclined individuals.

Three more questions from the mathematics subject practice test were added in order to insure that stereotype threat could occur for more highly math identified people. The percentage of correct answers for these questions was provided in the GRE mathematics subject test practice booklet. These percentages ranged from 52% correct to 73%. Because the test is divided into easy and hard questions, it was expected that a difference would exist between high math identified individuals and low math identified individuals regardless of stereotype threat conditions. It was thought that low math identified individuals would most likely do poorly on the harder questions. In order to make sure low math identified participants did not give up before completing the questions designed for them, the harder questions appeared last on the math test.

Implicit Measures (Implicit Association Tests). The Implicit Association Test (IAT) measures the strength of associations between different categories and provides an index of less conscious attitudes, identities, and stereotypes. Past research has used the computerized IAT, which measures the amount of time it takes for participants to categorize words into categories. Participants in the current study provided this information through the paper version of the IATs. Participants completed two critical IATs, one measuring associations between gender and the self (i.e. gender identification) and one measuring associations between math and the self (i.e. math identification). Before completing the Gender Identity IAT and the Math Identity IAT, participants also went through the practice IAT designed to measure attitudes about flowers and insects. This was included so that participants could get used to the test.

When completing the paper IAT, the participants categorized as many words as possible into different categories in 20 seconds. They marked an empty circle next to each word that indicated which category the word belonged to. Each category is made up of a word pair such as "flower/good". Participants should correctly categorize more words into a category (i.e. "flower/good") when the participant closely associates the words within that category. For example, if the participant implicitly thought of flowers as good, then they would categorize more words correctly when the category is "flower/good" then when it is "flower/bad".

Practice IAT (Flower/Insect Attitudes). In the flower/insect IAT, participants first classified words into the category "flower/good" with one response (i.e. marking a circle to the left of the word), and "insect/bad" with another response (i.e. marking a circle to the right of the word). In a separate section, category pairings switched, and participants classified words related to "flower/bad" with one response (i.e. left circle) and "insect/good" with the other (i.e. right circle). Participants were presented with a sheet of paper. The sheet of paper had two columns of words. Each word had a circle on the left of it and a circle on the right (i.e. one of two responses). Each circle column had a category heading. To make a response, participants marked the circle under the category heading that the word belonged in.

Participants were told to look at the category headings on top of the circle columns. The left category heading consisted of the words "flower" and "good" while the right category heading consisted of the words "insect" and "bad". The word lists were made up of words related to "flower", "insect", "bad", and "good". Participants were told that they were not deciding whether or not flowers and insects were good or bad, they were simply putting the flower words into the flower category, insect words into the insect category, bad words into the bad category and good words into the good category. For example: the word "lily" would go into the "flower/good" category because it is a flower word and the word "pleasant" would go in the "flower/good" category because it is a good word. Definitions of which words fit which categories were provided at the top of the page in case participants became confused. Words used for the flower/insect IAT were adapted from Greenwald et al., (1998). Full word lists can be found in Appendix A.

This first page represents one block of the paper IAT. Participants were given 20 seconds to correctly categorize as many words as possible in this first block. Participants were told not to stop and correct mistakes and that if they made it to the bottom of the first word column they must go on to the second until the 20 seconds were up.

After the first page (i.e. the first block), participants turned to the second page. This page was almost identical to the first page with two major differences. First, the word list was in a different order. Second, and most importantly, two of the words making up the category headings ("good" and "bad") switched sides. Instead of a "flower/good" category and an "insect/bad" category, this block had a "flower/bad" category and "insect/good" category. Participants again categorized as many words as possible in 20 seconds. If a participant categorized more words in one block than in the other, the participant's association should be stronger between the words that were paired up in the categories for that block. In the practice IAT, if a participant categorized more words correctly when "flower/good" was paired together (i.e. the first block), than when "flower/bad" was paired together (i.e. the second block), that participant had a stronger association between flower and good words than between flower and bad words. This is interpreted to mean that the participant had more positive attitudes towards flowers, which correspond with more negative attitudes towards insects.

Gender Identity IAT. The Gender Identity IAT was used to measure the strength of participants' identification with male relative to female. The Gender Identity IAT was presented in the same format as the practice IAT. In the first block, “male/self” appeared at the top of the left response column and “female/other” appeared at the top of the right response column. In the second block the words “self” and “other” switched so that “male/other” was the left category heading and “female/self” was the right category heading. The word list was made up of first names (i.e. “Amanda” and “Josh”), adapted from Greenwald et al., 1998) and pronouns (i.e. “me” and “them”). These pronouns have been used in a number of studies that have utilized the IAT (Greenwald et al., 1998; Lane et al., 2011; Nosek et al., 2002). Full word lists are included in Appendix A.

Math Identity IAT. The Math Identity IAT was used to measure the strength of participants’ identification with math relative to language. The Math Identity IAT was presented in the same format as the other two IATs. In the first block “math/self” was the left category heading and “language/other” was the right category heading. In the second block the words “self” and “other” switched so that “math/other” was the left category heading and “language/self” was the right category heading. The word list was made up of math related words (i.e. “calculations”), language related words (i.e. “grammar”) and the same pronouns used in the Gender IAT (i.e. “me” and “them”). Words that were categorized as “math” or “language” were adapted from Nosek et al. (2002). Full word lists are included in Appendix A.

Explicit Measures. Explicit measures took place after the IATs and were administered in the form of questionnaires.

Gender and Math Identification Scales. Participants completed explicit measures of gender and math identification by indicating the extent to which they agreed with a series of statements. These statements were items taken from Luhtanen and Crocker (1992)'s Collective Self-Esteem Scale in which participants indicate their level of agreement with statements on a 7 point scale (1= strongly disagree, 7= strongly agree). The Gender Identification Scale consisted of three items ("In general, being a man or a woman is an important part of my self-image.", "My gender is unimportant to the kind of person I am." and "Overall, my gender has very little to do with who I am."). The last two items of the Gender Identification Scale were reverse coded. The Math Identification Scale also consisted of three items ("In general, being good at math is an important part of my self-image.", "Success in math is unimportant to the kind of person I am.", and "Overall, math achievement has very little to do with who I am."). The last two items on the Math Identification Scale were reverse coded as well.

Three filler items were included in the questionnaire in order to guard against participants trying to discover the hypothesis of the study, or thinking too hard about the how their own math identification and gender identification might interact. Filler items were taken from the Contingencies of Self-Worth scale (Crocker, Luhtanen, Cooper & Bouvrette, 2003). Minor adaptations were made to filler items. For example the item "Having a moral code is an important part of my self-image." from the Contingencies of Self-Worth scale (Crocker et al., 2003) was prefixed with the phrase "In general," so that the filler items matched the wording used for gender and math identification items. Gender and math identification items appeared in mixed order in the questionnaire. There were three filler items ("In general, having a moral code is an important part of my self-image.", "My self-esteem would suffer if I did something unethical." and " Overall, my self-worth is based on others' approval."). Higher scores on the Gender Identification Scale were representative of a higher identification with the participant's gender. A test of internal reliability revealed that the items in the Gender Identification Scale were strongly correlated with one another (Cronbach's $\alpha = 0.76$). Similarly, higher scores on the Math Identification Scale were representative of a higher identification with mathematics. A test of internal reliability revealed that the items in the Math Identification Scale were strongly correlated with one another (Cronbach's $\alpha = 0.80$).

Typicality Scale. Participants also completed a questionnaire in which they rated character traits on how typical they were of male or female genders. This measured which traits were rated as typical of "female" according to the perception of these traits as relevant or irrelevant to math. Participants indicated the gender typicality of the words on a five-point Likert scale (1=Female, 5=Male). The Typicality Scale consisted of 15 traits. Four of these traits were previously shown to be perceived as female and "math-relevant" (i.e. "female math-relevant traits). Six were perceived as female and "math-irrelevant" (i.e. "female math-irrelevant traits") (Pronin et al., 2004). Finally, five of these traits had not yet been tested, but were included on an exploratory basis. These last items were hypothesized to be perceived as male and "math-relevant" (i.e. "hypothesized-male" traits). Male math-irrelevant items were not created given that Hypothesis 2 focuses on comparing women's perception of female stereotypes only. Furthermore, since the association between male and math is so strong, it may be the case that no male traits are perceived as truly "math-irrelevant".

Perceived femininity and math-relevancy of the stereotypically female traits were validated by previous research (Pronin et al., 2004). In one study, participants confirmed that the "female" words were, indeed, perceived as feminine. In a second study, new participants rated these words in answer to the question, "How much you think each of these characteristics would put a woman at risk for being negatively judged in quantitatively based fields and/or careers?" (Pronin et al., 2004). Therefore, the "female math-relevant" and "female math-irrelevant" traits were tested for perceived femininity and level of math relevancy.

Seeing as the items on the Typicality Scale ask participants to label these traits based on gender, it was important to make sure these traits were as unambiguous as possible. Therefore, the trait "having children" was not used in the current study because the liberal arts population is known for its acceptance of blended gender roles. Also, the trait "having children" could be interpreted as "the ability to give birth" which is unarguably a female trait. The item "leaving work to raise children" was changed to "provides childcare" for similar reasons. The "female math-relevant" item "fashionable" was also excluded, seeing as "wearing makeup" and being "artistic" (i.e. "female math-irrelevant" traits) may coincide with "fashionable" in popular culture. The "female math-relevant" traits in the current study were "sensitive", "nurturing", "shy" and "empathetic". The "female math-irrelevant" traits were "flirtatious", "emotional", "gossipy", "wears makeup", "artistic" and "provides childcare".

An individual score for type of trait was the average of all items on the Typicality Scale that made up a particular trait type. Therefore, each participant had three typicality scores, one for "female math-relevant" traits, one for "female math-irrelevant" traits and one for “hypothesized-male” traits. Internal reliability was calculated for each trait type (“female math-relevant”: Cronbach’s α= 0.62, “female math-irrelevant”: Cronbach’s α= 0.70, “hypothesized-male”: Cronbach’s α= 0.71). Items were strongly correlated with each other. The slightly weaker correlation found within the “female math –relevant” trait ratings may have been due to the small amount of items representing this type (i.e. only four).

On an exploratory basis, words and phrases hypothesized to be perceived as more “math-relevant” than the female items were also added to the Typicality Scale. These items were “perseverance”, “organized”, “logical”, “focused” and “problem-solver”. Because men are stereotyped as being good at math, it was also hypothesized that these words would be rated as male in the central study. A pilot study was conducted in order to test the math relevancy of these “hypothesized-male” items. A total of 18 participants (7 male, 9 female, and 2 identifying with neither gender, *M* age=20) completed a questionnaire that asked them think about stereotypes of individuals who are either good or bad at math. They then rated the extent to which each type of person fit with a trait on a five-point Likert scale. The same 15 words that were used in the main study were used in the pilot study. Higher numbers meant the trait more accurately described a person is stereotypically good at math. If men are perceived as more relevant to math than women, “hypothesized-male” traits should have received higher scores than both types of female traits.

Results of the pilot study revealed a significant difference between all three types of traits, $F(2,34) = 48.05$, $p < .0001$, , $\eta_p^2 = 0.84$. Planned contrasts showed that “hypothesized-male” traits were indeed rated as significantly more math relevant (*M*= 4.62, *SD*= 0.54) than the “female math-relevant” traits (*M*= 2.85, *SD*= 0.38), $t(17) = -11.288$, $p < .001$, $d = -$

2.66, and the "female math-irrelevant" traits (*M*= 2.45, *SD*= 0.66), $t(17)=-9.95$, $p < .001$, d= 2.36. The difference between "female math-relevant" traits and "female math-irrelevant" traits found in Pronin et al. (2004) was also replicated, $t(17)=3.26$, $p=.005$, d= 0.78.

Demographic Questionnaire. Participants provided their intended major, top three career choices, gender, age, year in school, race/ethnicity, country/region of citizenship and whether or not English was their first language. They also indicated how the study went, whether they noticed anything unusual in the study, what they thought the hypothesis was, whether the math test was difficult and whether or not their name was in the IAT stimulus list.

Procedure

Participants completed the experiment both individually and in groups of no more than five. The quietness of the testing room varied with the recruitment location. When recruiting in the campus center, there was a fair amount of muffled noise outside the room. Participants in the lab had a quieter atmosphere. Still, both areas were private and minimized distractions. After providing informed consent, the experimenter (Caucasian female) informed participants that they were about to take a math test and conducted either the threat or non-threat manipulation. Participants were randomly assigned to either the stereotype threat or non-threat condition. Participants were then told that the math test consisted of six multiple choice questions that they had a total of 10 minutes to finish. The experimenter also explained that the participant should circle their best guess if they were unsure of how to answer.

When they were finished with the math test, or when 10 minutes had passed, all participants completed the implicit identification measurements. The experimenter announced that they were about to complete a word categorization task. Instructions for the task were given and the experimenter timed each 20 second block using a cellphone stopwatch. Each participant first completed the Practice IAT, the Gender IAT and finally the Math IAT. Participants then completed the explicit identification questionnaire, the math-relevancy questionnaire and a series of demographic questions. They were debriefed on the goals of the study, the effect of stereotypes on math performance and how this differs according to gender. Lastly, participants were thanked for their time and compensated.

Results

Data Preparation

Data preparation for analyses using explicit measures was minimal. Three participants were excluded on the basis of suspicion of stereotype threat, leaving a total of 68 participants. If analyses involved gender as an independent variable an additional four participants who did not identify as either male or female were excluded from these analyses (N=64).

Data preparation for analyses using implicit measures controlled for error during the administration of the IATs. In order to make sure participants had correctly understood the instructions for the IAT, a series of exclusions were made. Participants were excluded if they received a score of below 8, marked greater than 20% of the answers incorrectly, and if English was not the participant's first language. In addition, one participant stopped in the middle of the IAT and another admitted to completing the task incorrectly. The number of participants varied depending on the type of implicit measure included in the analysis (i.e. math identification or gender identification).

The IAT was scored according to the methods laid out by Lemm, Lane, Sattler, Khan & Nosek (2008). More positive numbers represented higher implicit identification with “female” whereas more negative numbers represented higher implicit identification with “male”.

Stereotype Threat

Before conducting analyses that would test the primary hypotheses, it was necessary to determine whether or not stereotype threat had occurred. A 2(gender: male/female) x 2(condition: threat/non-threat) between subjects Analysis of Variance (ANOVA) was used to measure differences in the total number of correct answers on the math test. Due to experimenter error, the correct answer to question one did not appear on the math test. This question was taken out of analysis, leaving a total of five questions.

The ANOVA revealed no significant main effects of either gender, $F(1,60) = 0.002$, $p= 0.995$, *ns,* $\eta^2 < .001$ or condition, $F(1, 60) = 0.423$, $p= 0.518$, *ns,* $\eta^2 < .001$. However, there was a significant interaction between gender and condition such that the mean test score of women in the threat condition was lower ($M=1.88$, $SD= 0.21$) than women in the non-threat condition ($M= 2.23$, $SD= 0.19$), $F(1, 60) = 4.84$, $p= 0.03$, $\eta^2 =0.01$. Conversely mean test score of men was higher in the threat condition ($M= 2.39$, $SD= 0.24$) and lower in the non-threat condition ($M= 1.75$, $SD= 0.25$). Planned comparisons showed that the difference between women’s test scores in the threat condition versus the non-threat condition was not significant, $t(37) = 1.34$, $p= 0.20$, *ns*, $d= -0.44$. This result was the same for men. The comparison of men in the threat and non-threat conditions did not reaching significance, t(23)= 1.65, *p*= 0.12, *d*= 0.65. Even though planned comparisons were not perfectly in agreement with previous research (Ben-Zeev et al., 2005; Lesko & Corpus, 2006; Wout et al., 2008), the significant interaction states that differences between means are due to the combined effect of threat condition and gender. In other words, the effect of threat on math performance depended on gender (refer to Figure 3). Considering that the critical interaction was detected and the small effect of this interaction replicated previous research (Lesko & Corpus, 2006)[3], the sample size may have been too small to allow planned comparisons to detect significance (Threat Condition: 13 men/ 17 women, Non-Threat Condition: 12 men/ 22 women).

Furthermore, if the interaction between conditions had a real effect on test score, than male and female scores should have been the same when condition was not taken into consideration. This is exactly what was discovered. No difference was found between the average test score of men and women overall (Men: *M*= 2.08, *SD*= 1.00, Women: *M*= 2.08, *SD*= 0.81). This provides further evidence that planned comparisons did not reach full significance because of the small sample size and not because stereotype threat did not occur. These results were also replicated when suspicious participants were taken out of analysis. When asked what they thought the hypothesis of the study was, only 7 (10.3%) participants specifically mentioned the stereotype threat manipulation. The ANOVA was repeated leaving out suspicious participants with the same pattern of results. The

[3]Although reports of effect sizes are scarce, there is a possible pattern in previous research (Lesko & Corpus, 2006; Ben-Zeev et al., 2005; Wout et al., 2008) in which there is a small effect of stereotype interaction, but a large effect of planned comparison. This is expanded upon in the discussion.

interaction between condition and gender was still significant with means in the predicted direction, $F(1,54) = 4.50$, $p = 0.04$, $\eta^2 = 0.01$. This is not surprising, considering the inclusion of suspicious participants might have made it less likely to detect stereotype threat.

Interestingly, there was a trend in which men in the threat condition had higher test scores ($M = 2.39$, $SD = 0.24$) than men in the non-threat condition ($M = 1.75$, $SD = 0.25$) (refer to Figure 3). Again, this difference was non-significant. The significance value ($p = 0.12$) does suggest that this was a strong trend. For the reasons referred to above this trend is thought to be due to the small sample size and not to a weakness of effect. This means that the stereotype threat manipulation may have bolstered men's math scores. Most intriguing was that the effect size of men between condition was larger than the effect size of women between condition (Men, $d = 0.65$, women, $d = -0.44$). This, coupled with the fact that the planned comparison for women was non-significant, suggests that men were affected more strongly by stereotype threat, but in the opposite direction. Men's math performance was bolstered to a higher degree than women's math performance was obstructed.

Hypothesis 1: Gender Identity

The first hypothesis stated that explicit gender identification would be increased in women under stereotype threat, but not in women in the non-threat condition. Overall, women reported an average gender identification score of 4.84 ($SD = 1.36$) revealing that women do not report especially low or high gender identification, but that they lean slightly towards higher identification. Women's average gender identification score also did not significantly differ from men's score, $t(62) = 0.81$, $p = 0.420$, *ns,* $d = 0.21$. A t-test for independent means was conducted in order to examine Hypothesis 1 with condition as the independent variable and explicit gender identification as the dependent variable. Only women were used for this analysis (N=39).

The t-test comparing the explicit gender identity scores of women in the threat condition and women in the non-threat condition was non-significant, $t(37) = 0.959$, $p = 0.344$, *ns*, $d = 0.31$. Importantly the means were in the predicted direction with women under threat reporting stronger gender identity ($M = 5.08$, $SD = 1.15$) than women not under threat ($M = 4.65$, $SD = 1.51$). It cannot be concluded that women's gender identification was significantly impacted by being under stereotype threat. However, the small sample size increases the possibility that this difference could not be accurately detected in the current sample.

On an exploratory basis, analyses were conducted in order to find out whether or not the predicted patterns would emerge when implicit measurements were used. The overall breakdown of gender identification scores was unsurprising. Women associated words denoting "self" with words denoting "female" significantly more strongly than men , $t(43) = 4.68$, $p < .001$, $d = 1.48$ (Women: $M = 2.99$, $SD = 3.27$, Men: $M = -1.21$, $SD = -1.80$). This suggests that implicit female identification was stronger for participants who indicated they were female, than for participants who indicated they were male.

Again, only women were used for the main analysis. After IAT exclusions, the data of 29 women were used in the critical t-test. The critical t-test replicated results found with explicit measurements, revealing no significant difference between women in the threat condition ($M = 3.59$, $SD = 2.46$) and women in the non-threat condition ($M = 2.56$, $SD = 2.83$), $t(27) = 1.02$, $p = 0.318$, *ns*, $d = 0.38$, with means in the predicted direction (refer to Table 1 for a summary of both explicit and implicit means). Hypothesis 1 was therefore not fully supported, as both explicit and implicit measures of gender identification were not significantly increased by the stereotype threat manipulation. It is, however, notable that means were in the predicted direction on both explicit and implicit levels. This agreement between explicit and implicit measures makes it more likely that this pattern may be found with a larger sample size. In addition, the effect sizes of critical t-tests are medium. It is possible that these medium effects would have been detected if a larger sample size was used.

Hypothesis 2: Typicality Scale

The second hypothesis stated that women with high math identification who are under stereotype threat will judge "female math-irrelevant" traits as less descriptive of their gender. Lower scores on the Typicality Scale meant the participant rated words as more feminine while higher scores coincide with more masculine ratings. If the hypothesis is supported, women under threat who have high math identification should give higher ratings (i.e. a more masculine rating) to "female math-irrelevant" traits than women under threat with low math identification. The "female math-irrelevant" traits should also receive lower ratings than "female math-relevant traits" from highly math identified women. Since there were two measurements of math identification (i.e. explicit and implicit) two 2x2x3 mixed measure ANOVAs were used to test this hypothesis: 2(Condition: Threat/Non-Threat) x 2(Math Identity: High/Low) x 3(Trait Type: "female math-relevant"/"female math-irrelevant"/"hypothesized-male"). As in Hypothesis 1, men were excluded from these analyses. The data of 39 women were used for in the analysis using explicit measures, whereas the analysis using implicit measures consisted of 31 women.

The explicit mixed measures ANOVA used the Math Identity Scale to differentiate between women with high and low math identification. This differentiation was made with a median split procedure. Women whose scores on the Math Identity Scale were above the median (3.00) were considered high in explicit math identification (N=19) and women whose scores were below the median were considered low in explicit math identification (N=20).

Contrary to the hypothesis, the mixed measures ANOVA using explicit math identification did not reveal a significant 3-way interaction between condition, math identification and trait type. Mean ratings of traits were in the opposite direction than predicted. As shown in the left graph of Figure 4, highly math identified women in the threat condition rated "female math-irrelevant" traits as (non-significantly) more typical of females (*M*= 2.13, *SD*= 0.65) and women who were low in math identification rated "female math-irrelevant" traits as (non-significantly) more typical of males (*M*= 2.20, *SD*= 0.26). Again, this difference did not significantly dependent on math identification or threat condition. However, the interaction would have been in the opposite direction than predicted. The surprising ratings of "female math-irrelevant" traits as *more* typical of the female gender than "female math-relevant" traits was not significantly dependent on higher levels of math identification and stereotype threat.

A significant main effect of trait type was found, $F(2,70) = 22.31$, $p<.0001$, $\eta_p^2 = 0.35$. Planned contrasts showed that "hypothesized-male" traits (discovered to be perceived as the most math relevant traits in the pilot study) were rated as more typical of males than female traits. In other words, "male" traits were rated as more typically male (*M*= 2.85, *SD*= 0.52) than both "female math-irrelevant" traits (*M*= 2.14, *SD*= 0.42), *t*(38) = -5.49, *p*<.001, *d*= -0.89, and "female math-relevant" traits (*M*= 2.24, *SD*= 0.50), *t*(38) = -4.43, *p*<.001, *d*= -0.71 (Refer to the left graph of Figure 5). The difference between ratings of "female math-irrelevant" traits and "female math-relevant" trait was marginally significant, *t*(38) = -1.79, *p*= 0.08, *d*= -0.30. No other main effects were significant.

Results show that ratings of the different types of traits did not significantly depend on condition, math identification or trait type. Mean patterns were opposite of the direction predicted. This pattern is mirrored in the significant main effect of overall trait ratings. In sum, both highly math identified women and women overall rated "female math-irrelevant" traits as the most typically female. This trend was significant for the overall sample of women, but not (as hypothesized) for highly math identified women under threat

The second ANOVA differentiated between women with high and low math identification using the Math IAT. Again, this was accomplished using a median split. Women who had IAT scores above the median (-1.98) were considered high in implicit math identification (N=16) and women who had scores below the median were considered low in implicit math identification (N=15).

Similar to results using explicit measures, the implicit mixed measures ANOVA found the hypothesized 3-way interaction to be non-significant. Again, mean ratings of words were *not* in the predicted direction. As shown by the right graph in Figure 4, highly math identified women in the threat condition rated "female math-irrelevant" traits as (non-significantly) more typical of females (*M*= 1.97, *SD*= 0.38) than women under threat who were low in math identification (*M*= 2.32, *SD*= 0.39). This difference did not depend on math identification or threat condition.

As in the previous test, the implicit mixed measures ANOVA yielded a significant main effect of trait type $F(2,54)= 16.37$, $p<.0001$, $\eta_p^2= 0.41$. Replicating results of explicit measures analyses, "male" traits were rated as more typically male (*M*= 2.85, *SD*= 0.57) than both "female math-irrelevant" traits (*M*= 2.14, *SD*= 0.44), $t(31) = 4.62$, $p<.001$, $d= 0.82$, and "female math-relevant" traits (*M*= 2.06, *SD*= 0.49), $t(31) = 0.318$, $p<.001$, $d= 0.72$ (Refer to the right graph of Figure 5). Unlike the results of the ANOVA using explicit measures of math identification, the difference between "female math-irrelevant" traits and "female math-relevant" traits was non-significant. Interestingly, these means were in the opposite direction of the previous ANOVA. Women rated the "female math-irrelevant" traits as (non-significantly) *less* typical of the female gender than "female math-relevant" traits. This difference was also slightly smaller than the marginally significant difference found in the previous ANOVA (*d*= -0.19 compared to *d*= -0.30). Since the main effect of trait type did not depend on implicit math identification scores, the only difference between the explicit and implicit measures was that the sample size in the implicit analysis was smaller. The seemingly contradictory pattern of "female math-irrelevant" and "female math-relevant" trait ratings in the second ANOVA was most likely due to this difference in sample sizes. There were no other significant main effects.

Contrary to the hypothesis, women's ratings of traits as typical of either male or female were not significantly affected by stereotype threat or their level of math identification. Furthermore, even if the differences were significant, the means showed an opposite trend than predicted. These results were the same regardless of how math identification was measured. Separate from these results, the overall sample of women rated "female math-irrelevant" traits as marginally more typical of female in the explicit measures ANOVA. Strangely this pattern was not replicated in the implicit measures ANOVA, with a non-significant difference between means that were in the opposite direction. This discrepancy was, again, thought to be due to a lower sample size.

Exploratory Analysis: Math Identity

Although it was not included as a goal of this study, questions were raised about the influence of stereotype threat on female math identification. It was informally predicted that math identity may be increased by stereotype threat. In order to test this prediction, the average math identity of women in the threat condition was compared to the average math identity of women in the non-threat condition. Two separate analyses were done. One analysis was completed with explicit measures of math identity and one with implicit measures.

Overall, women reported an average math identification score of 3.35 (SD= 1.40) revealing that women do not report especially low or high math identification, but do lean towards less identification. Women's reported math identification was not significantly different from men's math identification, $t(62)$= 0.67, p= 0.50, *ns,* d= 0.17. However, these means did *not* replicate previous research (Nosek et al., 2002; Nosek & Smyth, 2011), with women reporting surprisingly greater identification (M= 3.35, SD= 1.40) with math than men (M= 3.26, SD= 1.49). Again, this difference was not significant.

The first critical t-test compared the explicit math identity of women in the threat condition and women in the non-threat condition. This was significant, $t(37)= -2.35$, $p= 0.02$, $d= 0.76$. Interestingly, this significant difference was *not* in the predicted direction. Women in the threat condition reported less identification with math ($M= 2.78$, $SD= 1.15$) than women in the non-threat condition ($M= 3.79$, $SD= 1.44$).

The second critical t-test used implicit math identification. The IAT scores of 31 women were selected for analysis after exclusions. More positive numbers represent stronger identification with math while more negative numbers represent stronger identification with language. In agreement with previous research, women showed significantly less identification with math ($M= -1.41$, $SD= 2.51$) than men ($M= 1.13$, $SD= 3.88$), $t(25.45) = -2.50$, $p= 0.02$, d= -0.83.

The difference between implicit math identification for women under threat compared to women not under threat was marginally significant, $t(26.93) = -1.77$, $p= 0.09$, $d= -1.27$. This marginal trend partially replicated the analysis of the explicit math identification scores. The means were also *not* in the predicted direction (threat: $M= -2.25$, $SD= 1.54$, non-threat: $M= -0.81$, $SD= 2.93$) (refer to Table 2 for a summary of both implicit and explicit means). Women who were under the threat manipulation had stronger associations between female names and language related words than between female names and math related words. The prediction that stereotype threat would increase math identification was refuted by the current data. In fact, the opposite occurred. Stereotype threat significantly decreased explicit math identification in women. Stereotype threat also decreased implicit math identification in women to a marginally significant degree.

Discussion

Analyses revealed that stereotype threat did in fact occur in the current sample. Women who were made aware of the gender stereotype regarding math performed worse on the math test than women who were told information that discredited the stereotype. Analyses of Hypothesis 1 did not show that gender identity was significantly impacted by the stereotype threat manipulation on either an explicit or implicit level. However, means were in the predicted direction. Results did not support Hypothesis 2. Women high in math identification did not rate "female math-irrelevant" words as significantly more typical of the female gender than women low in math identification while under stereotype threat. This was true using both explicit and implicit measures of math identification. Exploratory analyses showed some unexpected, but significant, patterns regarding the impact of stereotype threat on math identification. Contrary to informal predictions, explicit math identification was decreased in women under the threat condition. This pattern was marginally significant for implicit math identification. Implications of these results and limitations of the current research are discussed below.

Gender Identity

According to Hypothesis 1, female gender identification should have been increased by stereotype threat, leading to larger explicit means and more positive implicit means for women in the threat condition. Although the difference was not significant, means were in the predicted direction both on the explicit and implicit levels. There are a few explanations for this outcome. One is, of course, that there really isn't a causal relationship between stereotype threat and gender identification. The other, is that the relationship does exist, but significance could not be detected because of the small sample size. Comparing the results of the current research to that of previous research will lead to clearer interpretations.

Out of the small pool of research which analyzed the relationship between stereotype threat and gender identity, there is evidence for and against the existence of this relationship (Kiefer & Sekaquaptewa, 2007; Wout et al., 2008). Previously, the study that had not found this causal relationship (Kiefer & Sekaquaptewa, 2007) did not emphasize gender in the stereotype manipulation procedure. The study reporting such a relationship *did* emphasize gender (Wout et al., 2008). Interestingly, gender was explicitly mentioned in the current study. This is surprising, since the causal relationship between stereotype threat and gender identification did not reach significance. Participants were clearly reminded that the gender stereotype existed and that men had previously performed better than women on the math test (i.e. threat) or that they had performed equally (i.e. non-threat). If the non-significant results of the current study are representative of reality, then stereotype threat does not affect gender identification. If this is the case even when gender is mentioned in the stereotype manipulation, then the type of stereotype manipulation used should have no effect on the relationship between gender identity and stereotype threat. This is contrary to research maintaining that gender identification is related to stereotype threat only when the manipulation makes gender salient (Wout et al., 2008). The non-significance of the current results, therefore, does not fit well with previous research.

Previous research that found no relationship between stereotype threat and gender identity had participants fill out measurements of gender identification before the stereotype threat manipulation. Since the current study had participants fill out these measures afterwards, results would suggest that both initial gender identification and resulting gender identification is unrelated to the strength of stereotype threat effects. In other words, if the non-significance of the current study is accepted, it would mean that gender identification is completely unrelated to stereotype threat in mathematics.

It is extremely hard to believe that gender identification has nothing to do with how much a woman is affected by stereotype threat. This is difficult to believe both because the stereotype is so connected to gender (Kiefer & Sekaquaptewa, 2007; Nosek, et al., 2002; Nosek & Smyth, 2011) and gender identification and math identification have an inverse relationship (Nosek et al., 2002; Park et al., 2011; Pronin et al., 2004; Wout, et al., 2008). It is, therefore, likely that the sample size was too small to detect significant differences between women under threat and women not under threat. Stereotype threat effects are usually small (Lesko & Corpus, 2006). A power analysis using G*Power 3.1 software showed that in order to have 80% power in the analyses for Hypothesis 1, the ideal sample size would be 788 participants. The current study had 71 participants not including exclusions. Since women were the only participants included in these analyses, this number was basically halved. Analysis using explicit measures had 39 women and analysis using implicit was left with 29. Needless to say, power was extremely low. Even with this low power, the difference between means was not small. There was a medium effect size for differences in both explicit (Cohen's *d*= 0.31) and implicit (Cohen's *d*= 0.38) gender identification. These interpretations of the current results should still be accepted with caution considering the difference is not significant. However, it is likely that a type two error has occurred.

Perceptions of "Female"

Perception of "Female" Unrelated to Gender Identity. The second hypothesis stated that women under threat who are high in math identification should have a different opinion of what trait types are typical of the female gender. Critically, this opinion should be different compared to women under threat who are low in math identification. More specifically, highly math identified women should have rated stereotypical female traits that are perceived to be "math-irrelevant" as less female than women who are low in math identification. It was, therefore, conceptually hypothesized that the combination of being highly math identified and under threat would cause women to perceive the female gender as more relevant to math.

Results showed no significant three-way interaction between word type, level of math identification and threat condition. There was also no significant main effect of threat condition or level of math identification. These results were the same in analyses using explicit math identification and implicit math identification. If these results reflect reality, it means that stereotype threat works without changing explicit understanding of the female gender as a whole[4]. If this is the case, the way women perceive their gender may not be related to differing stereotype threat effects. These results imply that there is also no

---

[4] Although different analyses used explicit or implicit measures of math ID, perceptions of the female gender were only measured explicitly. Trait ratings, therefore, speak to women's explicit understanding of the female gender.

difference between the way math identified and math unidentified women view gender while under threat. At first, it may seem as if this contradicts previous research which reveals an inverse relationship between gender and math identification (Nosek et al., 2002; Park et al., 2011; Pronin et al., 2004; Wout, et al., 2008). This literature suggests that the level of math identification a woman has should influence thoughts about personal gender identification. It could be, however, that women's perception of the female gender is not synonymous with personal gender identification. A highly math identified woman may be able to think of themselves as an exception to the gender stereotype, while still internalizing the idea that the female gender is not synonymous with math. This is an idea that could be expanded upon in future research.

There is also the possibility that the un-hypothesized patterns *are* significantly different from each other and that the small sample size affected the ability of analyses to detect statistically significant differences. The samples size was extremely problematic for Hypothesis 2. There were four group combinations being compared in the mixed measures ANOVAs. These were women under threat with high math identification, women under threat with low math identification, women not under threat with high math identification and finally women not under threat with low math identification. The number of participants in each of these groups ranged from 6 to 13 in the analysis using explicit math identification and 5 to 11 in the analysis using implicit measurements. As these numbers illustrate, compared groups were extremely small and unevenly dispersed. On the one hand, the group size problem could mean that differences between trait type ratings would have actually been significant had there been more participant data. On the other hand, the extremely small and uneven comparison groups may render any patterns found to be the result of messy data and inconclusive.

Interestingly, non-significant trends were *not* in the predicted direction. In the threat condition, highly explicit math identified women's average ratings of "female math-irrelevant" traits were (non-significantly) more typical of female than low explicit math identified women's average ratings. This pattern was the same for women with differing levels of implicit math identification. Although the non-significant interaction means that this difference does not significantly depend on math identification, it was still necessary to examine the difference more closely. To address this issue, a planned comparison was conducted comparing highly explicit math identified women in the threat condition to low explicit math identified women in the threat condition on their ratings of "female math-irrelevant words". The t-test was not significant with a very small effect size (Cohen's d= -0.14). The low effect size and the non-significant result suggest that a difference does not exist and would not exist even if the sample size was bigger. This lends support to the interpretation discussed above; that level of explicit math identification in women under threat does not change the way women view their gender in terms of math-relevancy.

Planned comparisons using implicit math identification however, found a slight non-significant trend when comparing these two groups, $t(11) = -1.60$, $p = 0.14$, $d = -0.89$. The large effect size makes this non-significant difference more likely to be representative of real differences in perception than planned comparisons using explicit measures. The significance value may, in this case, have been due to sample size problems. Therefore, the non-hypothesized trend of "female math-irrelevant" traits being rated as more feminine by women under threat who are high in math identity is more likely to be the case if math identification is measured implicitly. These results suggest that this slight non-significant trend may have been significant at the implicit level with a larger sample size. If a replication with a larger sample size yielded such results, it would suggest (1) that this difference in perception of female is influenced by implicit math identification and therefore the processes involved may be under conscious awareness, and (2) that the more a threatened woman identifies with math, the more likely they are to view females as "math-irrelevant".

In the explicit analysis, regardless of math identification or threat condition, women similarly rated "female math-irrelevant" traits as the most feminine. Overall, women rated "hypothesized-male" words as the most male, "female math-relevant" words as in between the other two traits. While the difference between ratings of female and male traits were significant in planned contrasts, the difference between type of female traits was marginally significant with a medium effect size (Cohen's d = 0.30). On the implicit level this difference was in an opposite direction with "female math-irrelevant" words rated as second most male. However this difference was even smaller than when explicit measures were used (Cohen's d = -0.19) and non-significant. The failure of implicit analysis to replicate the marginal trend found in the explicit analysis is most likely due to the fact that the implicit analysis included fewer participants. The implicit analysis was so named because it used implicit math identification. Since math identification was not a comparison factor in this main effect of trait type, implicit and explicit main effect tests only differed in sample size. Regardless of math identification and threat condition women view the female gender as typically “math-irrelevant”.

Interestingly, whether or not the critical interaction between threat condition, math identification and trait type was significant, results can be interpreted in a similar fashion. Either math identification does not affect trait ratings, or identification with math leads to viewing female as more “math-irrelevant”. This trend in means for women divided by math identification does not differ from the trend of the overall female sample. Regardless of the significance of results, it seems that increased personal identification with math did not affect women’s perception of “female”. Since math identification and gender identification are linked by the stereotype, it would also seem that personal gender identification is also independent of women’s perception of “female”.

Figures 4 and 5 illustrate this point nicely. Most importantly, the left graph in Figure 5 shows that “math-irrelevant” traits were rated most feminine by the overall sample of women. Although results should be interpreted with caution, it was determined that this trend is most likely representative of reality. It is clear from Figure 4 that the pattern of average trait ratings in both graphs (non-significantly) mirror this pattern especially for highly math identified women. If women under stereotype threat who are high in math identity do indeed view female as significantly more “math-irrelevant”, this would not differentiate from the viewpoint of women in the general population. This again supports the idea that level of math identification may not change the way the female gender is viewed. Bear in mind, however, that a larger sample size would be needed in order to make more definitive conclusions.

Typicality Scale. Unrelated to the specific hypothesis, these results have provided a better picture of how traits are viewed according to gender and relevancy to mathematics. Previous research had divided traits into two groups. These groups were feminine words that were negatively stereotyped with math (i.e. "female math-irrelevant") and feminine words that were positively stereotyped with math (i.e. "female math-relevant") (Pronin et al., 2004). To maintain consistency, these category names were also used in the current research. However, including "hypothesized-male" traits has provided a better understanding of how these traits are perceived and why these category names might be misleading. In the pilot study, "hypothesized-male" traits were perceived to be significantly more descriptive of someone who is stereotypically good at math. Furthermore, the main results showed that these same words were rated as significantly more male than both types of female words. These effects were large, ranging from d= -0.71 to d=0.89. Therefore, the same traits that were perceived as typically male were also perceived as more relevant to math than even the "female math-relevant" traits. This suggests that "female math-relevant" traits are not particularly relevant to math after-all and are math-neutral at best. Adopting this terminology may make for clearer future research.

There were also limitations of the Typicality Scale that should be addressed. Research investigating racial stereotypes concerning Black Americans focused on the importance of framing questionnaire instructions clearly (Devine & Elliot, 1995). Devine and Elliot (1995) revisited the results of the Princeton Trilogy studies, which were conducted in 1933, 1951 and 1969. These studies found an apparent decrease in stereotyping in Princeton students by recording participants' selections of words that typically described a Black person (Gilbert, 1951; Katz & Braly, 1933; Karlins, Coffman & Walters, 1969; as cited in Devine & Elliot, 1995). However, findings of Devine and Elliot (1995) indicated that this decrease reflected lessened negative personal beliefs towards Black people, but that knowledge of the racial stereotype itself had not decreased. It was concluded that misleading results were due to unclear instructions that did not differentiate between personal beliefs and knowledge of stereotypes. Similarly, the Typicality Scale did not specify whether participants should have reported personal beliefs or knowledge of the stereotype. Indeed, some participants expressed confusion about instructions during testing. These participants were told to report their personal beliefs, but it is not clear that all participants did so. This may have led to noisy data, in which the typicality scale represented a mix of stereotype knowledge and personal belief.

Another limitation to the Typicality Scale is that it measured these viewpoints explicitly. Implicit and explicit measurements can both be beneficial methods of understanding attitudes. However, when these attitudes are held about socially sensitive subjects, implicit measurements have better predicted behavior (Greenwald et al., 2009). It is possible that participants did not record their true beliefs out of concern of appearing sexist. Participants may also have implicit associations about women that they are not consciously aware of and, therefore, unable to report. Future research striving to understand how women truly understand their own gender may benefit from the use of implicit measurements.

Math Identity

Some predictions were made about the possible effects of stereotype threat on math identification. As with gender identification, it was predicted that math identification would be increased by stereotype threat. These predictions were tested using both explicit math identification and implicit math identification. A significant difference was found between the explicit math identification of women in the threat condition and women in the non-threat condition. This difference reached marginal significance in the analysis using implicit measures. Although these differences were significant, the means were in the opposite direction than predicted. Instead of an increase in math identification, women under stereotype threat showed a decrease in math identification as compared to women not under threat.

The results of the current research are exciting as it shows that stereotype threat can decrease math identification in women. Since previous research had measured math identification before the stereotype threat manipulation took place (Steinberg et al., 2012), it could not be determined whether stereotype threat had a causal effect. The current research measured math identification after the manipulation, making it possible to conclude causality. The causal relationship is especially strong because it is true on both explicit and implicit levels. Merging results of previous and current research, it can be concluded that initial high levels of math identification causes stronger stereotype threat effects (Steinberg et al., 2012), while the stereotype threat manipulation itself causes math identification to decrease.

A continuity problem existed when it was predicted that stereotype threat would increase math identification. In order to deal with this continuity problem it was further hypothesized that women would have different perceptions of gender based on their math identification (Hypothesis 2). This would allow high math and high gender identification to exist at the same time in women under threat. The results of the exploratory analysis create a different picture (refer to Figure 6), in which stereotype threat increases gender identification and decreases math identification. If the non-significance of the relationship between stereotype threat and gender identification (Hypothesis 1) is due to the small sample size, than these exploratory results would fit in perfectly with the idea that math identification and gender identification are inversely related even when women are under threat.

As was already mentioned, there is evidence that this inverse relationship exists in the attitudes and identities of both genders (Nosek & Greenwald, 2002; Nosek & Smyth, 2011). This pattern was partially replicated in the current data. Men's reported gender identification was significantly positively correlated with reported math identification (r= 0.42, n=25, *p*= 0.04), while a negative correlation between these two identifications was non-significant, but strong in women (r= -0.25, n=39, *p*= 0.12).[5] If gender identification really is increased by stereotype threat then it would make sense that math identification is decreased by stereotype threat, erasing the original continuity problem.

Combining these exploratory results with interpretations of Hypothesis 2 analyses provides further evidence that personal identity and perception of gender are separate constructs. Interpretations of Hypothesis 2's results maintain that women's perception of the female gender exists independently of their math identification or gender identification.

[5] These correlations were not performed with implicit measures. However, considering patterns of implicit and explicit math and gender identification mirrored each other in this study, it can be said with reasonable confidence that the implicit correlations would show the same pattern.

If this is the case, then stereotype threat increases gender identification and decreases math identification without changing how women view the female gender. In other words, stereotype threat changes internal identification, but not the perception of overall gender norms. According to this model, a woman who is math identified most likely sees herself as an exception to the norm, not as an example of why the norm is invalid.

Stereotype Threat

Is The Bard Sample Unique?. One of the most captivating outcomes of this study is that the stereotype threat effect was replicated. There was always the possibility that the Bard sample, being a population thought to have generally liberal viewpoints towards gender, would not be susceptible to stereotype threat. Indeed, when questioned about explicit gender identification, participants reported an average score of 4.64 (*SD*= 1.43). This suggests that, on average, participants did *not* report that their gender was especially important to themselves. The IAT, however, revealed a slightly different story. On average, women reported higher identification with female and men reported higher identification with male. In previous research implicit gender identification has either shown the same pattern (Nosek et al., 2002) or no difference between genders (Lane et al., 2011). Unlike the current study, previous research also showed that explicit and implicit gender identification showed similar patterns (Lane et al., 2011, Nosek et al., 2002).Therefore, while Bard participants may not be unique in the way they are influenced by the gender stereotype, there is a surprising discrepancy between what kind of gender identification they explicitly report and implicitly reveal. Bard participants may hold egalitarian viewpoints in their conscious mind, while adhering to typical gender roles on a less-conscious level. Most intriguingly, Bard's generally egalitarian viewpoints about gender do not make them immune to stereotype threat. This is important seeing as people who are liberal in their views about gender may feel that they are above the influence of stereotypes. It is possible that ignoring the inevitable effect of stereotypes on behavior would actually increase the possibility of being affected by stereotype threat anxiety because there is no explanation as to where the anxiety comes from. Further research would be necessary to explore this particular hypothesis.

There may be a simpler explanation as to why implicit and explicit materials sometimes measured identification differently. Besides the fact that one measured conscious attitudes while the other tapped into less conscious ones, the explicit questionnaire could have been measuring the importance of a specific domain in isolation whereas the IAT compared two categories. In terms of gender identification, the IAT measures the association of self and "male" in relation to the association between self and "female". On the other hand, explicit measures asked participants to think about their own gender, and only that gender, in relation to the self. It may be that discrepancies between implicit and explicit analyses are due to this difference in measurement design. This does not change comparisons between the Bard sample and samples in previous research however, as explicit and implicit measures of gender identification were taken from previous research (Lane et al., 2011; Nosek et al., 2002; Nosek & Smyth, 2011).

At first, it seemed that a similar divergence between implicit and explicit measures of math identification was found in the Bard sample. Consistent with previous research (Nosek et al., 2002; Nosek & Smyth, 2011) women in the overall sample were significantly less implicitly identified with math than men. This effect size was not only significant, but large (*d*= -0.83). This pattern, however, was not replicated with explicit measures. The difference was non-significant, on top of which, women reported surprisingly greater explicit identification with math than men did. Although a slight majority of the math and science majors were women (12 out of 23), math identification between these groups was not significantly different from each other in a 2 (Gender: male/female) by 2 (Major: science or math/other) ANOVA. Therefore, even if the non-significance of this trend was due to the small sample size, it could not be explained by a female math major majority. The trend, however, was both small (*d*=0.17) and non-significant, suggesting that an effect did not occur. This would support previous research showing an agreement between explicit and implicit math identification (Nosek et al., 2002; Nosek & Smyth, 2011).

This interpretation of results suggests that Bard students do not differ from other population as far as math identification is concerned. In light of the current study's small sample size more research is needed to confirm or disconfirm the non-significant trend in which women reported surprisingly more explicit math identification. It is also important to note that the current research did not specifically compare Bard with outside samples. Research that highlights this type of comparison as a main goal is extremely necessary in understanding whether the sample was unique in their conscious reports of identity.

Disagreement Between Analyses. There was a puzzling discrepancy in analyses between the ANOVA and planned comparison results. The ANOVA showed a significant interaction between gender and condition. This means that differences in test scores depended on both participant's gender and condition. However, planned comparisons (analyzed with independent t-tests) did not reach significance. It can be concluded, however, that this was due to the small sample size. This is particularly likely seeing as there were no difference between male and female performance when condition was taken out of the equation (*d*= 0.00).

A previous report of a similarly small interaction effect adds to the evidence of stereotype threat (Lesko & Corpus, 2006)[6] . The effect size for the interaction reported by Lesko & Corpus was small (η^2= 0.01). However, the planned contrast had a fairly large effect size (η^2= 0.10). This is consistent with surprisingly large effects sizes of stereotype threat articles using t-tests as their main analysis (Ben-Zeev et al., 2005; Wout et al., 2008). Although results of the current study did not have significant t-tests, they show the same pattern of effect sizes. T-tests used as planned comparisons revealed a moderate (non-significant) effect size when comparing women in the threat condition to those in the non-threat condition (*d*= -0.44) as well as a large (non-significant) effect size when comparing men in the threat condition to those in the non-threat condition (*d*=0.65). Consistent with previous research, the same data revealed a small (significant) interaction effect (η^2=0.01). It may be that the effect of one variable depending on another (i.e. the interaction) is generally small, but comparing different pairs of groups (i.e. planned comparisons) leads to larger effects. Since the effect size and significance level of the current interaction line up with that of previous studies, than the non-significance of planned comparisons must have been due to the small sample.

[6] Unfortunately, the amount of comparison effect sizes is small seeing as effect sizes of stereotype threat interactions were not reported in much of the reviewed stereotype research.

The ability of the math test to adequately measure math performance may also have contributed to the non-significance of these planned comparisons. It could be that the math test in the current study was not long enough to adequately measure performance. The reviewed research in which t-tests were used for main analyses did not provide detailed descriptions of math tests. It is plain, however, that these test were longer than the one used in the current study. One test was 20 questions long and asked participants to identify the correct way to solve word problems, not requiring the problem to actually be solved (Wout et al., 2008). The other used another math test of 20 questions in which participants were required to solve problems (Ben-Zeev et al., 2005). The article that originally published evidence of stereotype threat in female math performance had a test with at least 11 questions (Spencer et al., 1999). The math test used in the current study had only five questions. Perhaps if the math tests included more items, the results of planned comparisons would have more closely replicated that of previous research.

Male Math Performance. Effect sizes also show a possible boost in mathematics performance for men. Although this boost can be observed in graphs within previously reviewed research articles, it is often not reported or concluded to be non-significant (Ben-Zeev et al., 2005; Lesko & Corpus, 2006; Martens et al., 2005; Spencer et al., 1999). The current results similarly show a large non-significant effect of condition on men's math performance. However, considering the non-significance of the t-tests is thought to be due to methodological problems in the current study, the effect size of this analysis may be more informative than the significance level. The boost in men's performance while under threat is a larger effect than the depression in women's math scores. This could be the case because male participants in this sample were especially likely to react positively to being told they were superior at math. It also could be that women were not especially likely to react negatively to being told they were inferior at math. In other words, although women were affected by stereotype threat, they may not have been as strongly affected as those in previous research. Although this is speculative, it would be exciting if explicit egalitarianism of the Bard sample regarding gender roles was enough to make stereotype threat effects slightly weaker, even if not completely eradicated. This is another topic for future research.

General Limitations

Testing Environment. Unfortunately, the testing environment was not ideal. In order to collect as much data as possible in a short amount of time, the majority of participants were tested in a private room in the campus center building. This room, however, was not soundproof. Music and conversation was audible from inside the room. An open window created other distractions (i.e. noise, one person walking through the window to get into the building). The rest of the participants were tested in a quiet lab, with little to no distractions. In addition, it was practical to test multiple students at once. This created situations in which some participants were being tested with friends or with gendered groups that may have lessened or amplified the stereotype threat effects (i.e. two men and one woman). If this study was repeated, it would be ideal to test participants individually and in a lab setting.

Overlap Participants. The second issue is that participants from the current study could have also participated in a similar study. Two senior projects were being recruited for in the same areas. There was no attempt made to recruit different people for different studies. Both studies used a math test that had at least one repetition problem. The other study also had a priming component, designed to change math performance. Nine out of 71 people participated in both studies. It is, however, fairly unlikely that people who participated in both studies had their behavior affected in a way that would change results.

First of all, participants were not run in experiments on the same days. Participants who may have been primed the day before, had most likely lost the priming effects by the time they participated in the study reported here. It is also unlikely that sharing a question on the math test influenced scores very much. It would seem that sharing a question on the math test would boost math scores for participants who had done both studies. In the current study, however, the average score was a failing grade. If a few scores were boosted, it was not enough to overly affect the mean score. In any case, the only way this would have affected the data is if all overlapping participants were one gender or in one of the threat conditions. Although the majority of the overlapping participants were female (6 out of 9), a boosted scores for women would have hindered stereotype threat effects that were revealed in results. Unfortunately, there is no way of knowing what condition these participants were in. Even without this knowledge it seems that results should not have been affected by this overlap.

IAT Order. Another limitation that should be mentioned is that the IATs were presented in the same order for each participant. In a perfect version of the study design, the flower/insect IATs would have remained at the beginning of the packet while the gender and math IATs would have been presented in random order. The specific hypotheses were most likely not affected by this lack of randomization. This is because gender identification and math identification were not being compared with each other. Only their individual relationships with stereotype threat were being compared. Furthermore, patterns of implicit identification based on gender in the current study mirror results of previous research with women identifying more strongly with female and less strongly with math than men.

Conclusions

The most conclusive discoveries revealed by the current research center around stereotype threat and math identification. Stereotype threat occurred in the current sample, even though the sample was small and participants reported generally egalitarian viewpoints towards gender identification. This replication shows that these effects are prevalent and can occur for participants who may have more fluid ideas about gender than the general population. Contrary to predictions, stereotype threat caused a decrease in women's math identification.

A trend of gender identification was also found in which gender identification was non-significantly increased by stereotype threat. Considering the size of the gender identification effects and the agreement found between implicit and explicit measurements, it was concluded that this trend may be significant with a larger sample size. Overall, results suggest that an inverse relationship is present between women's gender identification and math identification even when under stereotype threat. Furthermore, results of Hypothesis 2 analyses suggest that this separation of identities does not change women's perception of what makes up the female gender.

These conclusions nicely compliment current research on identity separation and stereotype threat. According to the current research, “math-irrelevant” traits tend to be rated as most typical of females. It follows that identification with female means dis-identification with math. This separation has personal significance to women who most likely strive to be anti-stereotypic examples of their gender (i.e. highly math identified, successful in the workplace). Highly math identified women under stereotype threat in mathematics were less likely to report that “female math-irrelevant” traits (i.e. gossipy or flirtatious) described themselves (Pronin et al., 2004). Female participants also showed a separation of female and male identities in the face of stereotype threat in the workplace (von Hippel, Walsh & Zouroudis, 2011). This reveals that personal identity is shaped by stereotype threat, if not permanently, at least while under threat itself. When the stereotype is based on gender, whether the basis is success in the workplace or in math, separation of personal identification may be inevitable.

Research about stereotype threat and identity separation suggests that women who are successful at overcoming stereotype threat, do not necessarily reconcile the stereotype. Instead, they may learn to identify with male and female as it suits them. Imagine being asked to embody a masculine or feminine physicality. This means you would be encouraged to move as stereotypically male or female as possible. Although the movements in this exercise would be exaggerated, it serves to illustrate how polarized gender stereotypes can be. Now, think about how many times a gendered movement may have been advantageous for you in different situations. For example: when leading a discussion, it could be helpful to embody a powerful, assertive stereotypically masculine body, whereas when talking your way out of speeding ticket it may be more helpful to embody a helpless, harmless, stereotypically feminine body. These examples are, of course, extremely anecdotal, but it serves to embellish research findings fairly well. I argue that with these bodily switches, comes psychological switches. These switches are most likely micro, but should still be able to affect behavior, such as performance on a math test.

In support of the idea that people can switch between gender identities depending on the situation, women in college who were given a stereotype manipulation mentioning gender performed worse on a math test than women who were given a manipulation mentioning college students (Rydell, Beilock & McConnell, 2009). This means that the stereotype threat effects can be buffered by female college students being reminded of their college identity instead of their gender. Furthermore, female college students who were given a stereotype threat manipulation highlighting both gender and college identities did not show the same detrimental effects on performance as women who were just given a manipulation about gender (Rydell et al., 2009). This indicates that women temporarily chose the identity that allowed them to succeed when this identity was made available to them. [7]

In order to combat stereotype threat, it may be necessary to make this information common knowledge. If stereotype threat *does* cause an increase in gender identification, this would be detrimental to female math performance. In this situation, stereotype threat would cause women to choose an identity that, unfortunately, does not allow them to succeed as well. It also seems that the woman must be reminded of a helpful identity (i.e.

[7] It is most likely that men choose different identities that allow them to succeed as well. However, since the gender/math stereotype is fairly positive for men, it would be unnecessary for men to separate identity in this particular domain.

being a college student/ being more masculine/ being a math person) in order to have better scores on a math test.

Advocating conscious micro changes in identity may have a backlash effect in groups that are negatively stereotyped with math performance and whose group identity is largely unchangeable on a situational basis (i.e. whether or not someone is a college student). The degree to which one feels especially masculine or feminine in any given situation is arguably more changeable than whether or not one is a college student. Bearing this in mind, it could be very helpful for women to embrace the idea that gender identity can be manipulated as a tool for better math performance. In other words, even if a person strongly identifies with female, they can set aside that identification before taking a math test. Perhaps if this message was reserved for women only, a backlash effect would be minimized

Of course, this would be a very temporary solution that may work better or worse on an individual basis. A better, more long, term solution would be to try to change the stereotype that women are bad at math. This is exactly what toy companies like Goldie Blox ™ is trying to do by designing and advertising toys geared towards girls and engineering. However, with focus on the youth, a whole population of adult women who grew up with very gendered toys may not benefit from such initiatives. Perhaps learning to use these fairly established gender roles to their advantage is a better way of reducing adult female anxiety about math than asking them to change their entire viewpoints about an ingrained stereotype. As is suggested by the current research, perceptions about gender are more difficult to influence than personal identification. Giving women tools to decrease anxiety, instead of asking them to rearrange their whole system of beliefs may be a more effective way to change stereotype threat outcomes in the short term. Overtime, this could work with strategies like that of Goldie Blox™ to change gender stereotypes not just for the younger generation, but for older generations as well.

References

Barber, S. J., Mather, M., (2013). Stereotype threat can both enhance and impair older adults' memory.*Psychological Science*, *24*(12), 2522-2529. doi: 10.1177/0956797613497023

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*(2), 174-181. doi:10.1016/j.jesp.2003.11.007

Ben-Zeev, A., Dennehy, T. C., Sackman, R., Olide, A., & Berger, C. C. (2011). Flirting with threat: Social identity and the perils of the female communality prescription. *Journal of Experimental Social Psychology*, *47*(6), 1308-1311. doi:10.1016/j.jesp.2011.05.016

Cohen, G. L., & Garcia, J. (2005). "I am us": Negative stereotypes as collective threats. *Journal of Personality and Social Psychology*, *89*(4), 566-582. doi: 10.1037/0022-3514.89.4.566

Crocker, J., Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608-630. doi: 10.1037/0033-295X.96.4.608

Crocker, J., Luhtanen, R. K., Cooper, M. L., Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology. 85*(5), 894-908. doi:10.1037/0022-3514.85.5.894.

Devine, P. G., Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychology Bulletin, 21*(11), 1139-1150. doi:10.1177/01461672952111002

Educational Testing Service. (2007). *Practicing to Take the GRE General Test 10th Edition.* US: Educational Testing Service.

Educational Testing Service. (2008). *Graduate Record Examinations Mathematics Test Practice Book.* Retreived from http://www.ets.org/s/gre/pdf/practice_book_math.pdf

Goldie Blox™. (2014). Retrieved from http://www.goldieblox.com/

Goldman, E. G. (2012). Lipstick and labcoats: Undergraduate women's gender negotiation in stem fields. *NASPA Journal About Women in Higher Education*, *5*(2), 115-140. doi:10.1515/njawhe-2012-1098

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test.. *Journal of Personality and Social Psychology*, *74*, 1464-1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.. *Journal of Personality and Social Psychology*, *97*, 17-41.doi: 10.1037/a0015575

von Hippel, C., Walsh, A. M., Souroudis, A. (2011). Identity separation in response to stereotype threat. *Social Psychological and Personality Science*, *2*(3), 317-324. doi: 10.1177/1948550610390391

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, *2*(1), 204-222. doi: 10.1111/j.1751-9004.2007.00066.x

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology* , *43*(5), 825-832. doi: doi:10.1016/j.jesp.2006.08.004

Lane, K. A., Goh, J. X., & Driver-Linn, E. (2011). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, *66*(3-4), 220-234. doi:10.1007/s11199-011-0036-z

Lemm, K. M., Lane, K. A., Sattler, D. N., Khan, S. R., Nosek, B. A.(2008). Assessing implicit cognitions with a paper-format implicit association test. In Morrison, M. A. (Ed.) & Morrison, T. G. (E.d.), *The psychology of modern prejudice* (pp. 123-146). Hauppauge, NY, US: Nova Science Publishers.

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. *Sex Roles*, *54*(1-2), 113-125. doi: 10.1007/s11199-005-9973-2

Luhtanen, R., Crocker, J.(1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*(3), 302-318. doi:10.1177/0146167292183006

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: Threat effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*(2), 236-243. doi:10.1016/j.jesp.2005.04.010

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41) , 16474-16479. doi: 10.1073/pnas.1211286109

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math=male, me=female, therefore math=/me.*Journal of Personality and Social Psychology*,*83*(1), 44-59. doi:10.1037//0022-3514.83.1.44

Nosek, B. A., & Smyth, F. L. (2011). Implicit Social Cognitions Predict Sex Differences in Math Engagement and Achievement. *American Educational Research Journal*, *48*(5) 1125-1156. doi: 10.3102/0002831211410683

Park, L. E., Young, A. F., Troisi, J. D., & Pinkus, R. T. (2011). Effects of everyday romantic goal pursuit on women's attitudes toward math and science.*Personality and Social Psychology Bulletin*,*37*(9), 1259-1273. doi:10.1177/0146167211408436

Pollack, E. (2013, October 3). Why are there still so few women in science?. *New York Times.* Retrieved from http://www.nytimes.com/2013/10/06/magazine/why-are-there-still-so-few-women-in-science.html?_r=0

Pronin, E., Claude, S. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, *40*(2), 152-168. doi:10.1016/S0022-1031(03)00088-X

Rydell, R. J., McConnell, A. R., Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96*(5), 949-966. doi: 10.1037/a0014846

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

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test-performance of African-Americans. . *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629. doi: 10.1037/0003-066X.52.6.613

Steinberg, J. R., Okun, M. A., & Aiken, L. S. (2012). Calculus gpa and math identification as moderators of stereotype threat in highly persistent women. *Basic and Applied Social Psychology, 34*(6), 534-543. doi: 10.1080/01973533.2012.727319

Stone, J., Lynch, C. I., Sjomeling, M. Darley, J. M. (1999). Stereotype threat effect on Black and White athletic performance. *Journal of Personality and Social Psychology, 77*(6), 1213-1227, doi: 10.1037/0022-3514.77.6.1213

U.S. National Science Foundation. (2014a). *Science and engineering indicators*. *Average NAEP mathematics scores of students in grades 4 and 8, by student and school characteristics: 1990-2011.* Retrieved February, 2014, from http://www.nsf.gov/statistics/seind14/content/chapter-1/at01-02.pdf

U.S. National Science Foundation. (2014b). *Science and engineering indicators*. *Freshmen intending S&E major, by field, sex, and race or ethnicity: 1998-2012.* Retrieved February, 2014, from http://www.nsf.gov/statistics/seind14/content/chapter-2/at02-16.pdf

U.S. National Science Foundation. (2014c). *Science and engineering indicators*. *Earned bachelor's degrees, by sex and field: 2000-11.* Retrieved February, 2014, from http://www.nsf.gov/statistics/seind14/content/chapter-2/at02-17.pdf

U.S. National Science Foundation. (2014d). *Science and engineering indicators*. *S&E graduate enrollment, by sex and field: 2000-11.* Retrieved February, 2014, from http://www.nsf.gov/statistics/seind14/content/chapter-2/at02-24.pdf

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat . *Journal of Experimental Social Psychology*, *44*(3), 792-799. doi:10.1016/j.jesp.2007.07.005

Zhang, S., Schmader, T., Hall, W.M. (2012): L'eggo My Ego: Reducing the Gender Gap in Math by Unlinking the Self from Performance. *Self and Identity*,*12*(4), 400-412. doi:10.1080/15298868.2012.687012

Table 1. Means and Standard Deviations for Female Participants Gender ID

	Threat		Non-Threat		*p*	*d*
	M	SD	M	SD		
Explicit Gender ID	5.08	1.15	4.65	1.51	0.344	0.31
Implicit Gender ID	3.59	2.83	2.56	2.83	0.318	0.38

Note. Differences are not significant.

Table 2. Means and Standard Deviations for Female Participants Math ID

	Threat		Non-Threat		*p*	*d*
	M	SD	M	SD		
Explicit Math ID	2.78	1.15	3.79	1.44	0.02	0.76
Implicit Math ID	-2.25	1.54	-0.81	2.93	0.09	-1.27

Note. Differences are significant for explicit measures and marginally significant for implicit measures.

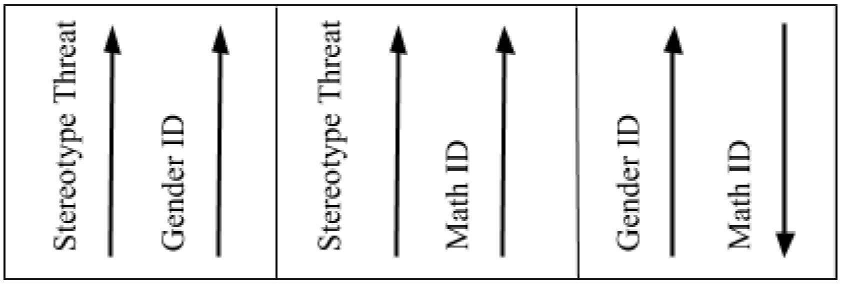


Figure 1. Hypothesized relationship between gender ID, math ID and mathematics-related stereotype threat in women.

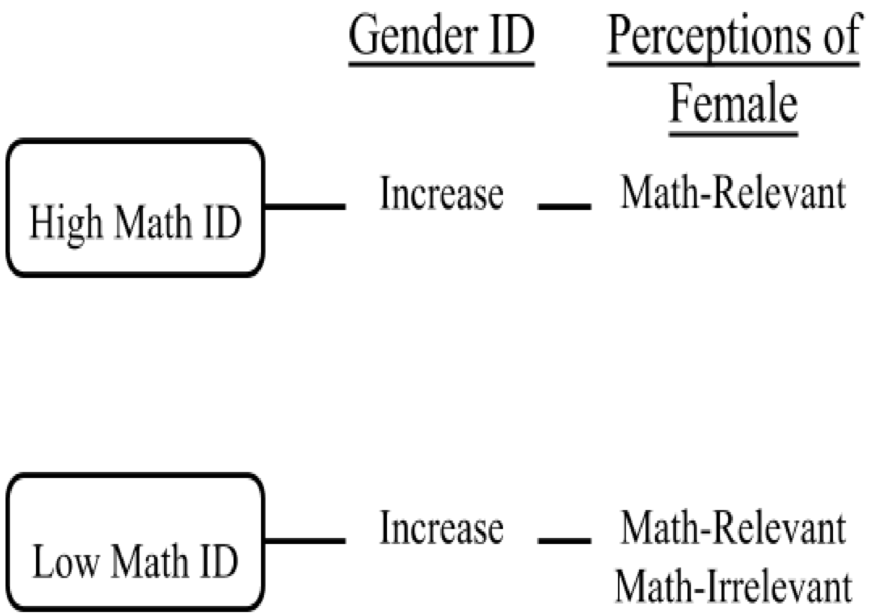


Figure 2. Hypothesized change in women's perception of "female" after a stereotype threat manipulation according to initial level of math identification.

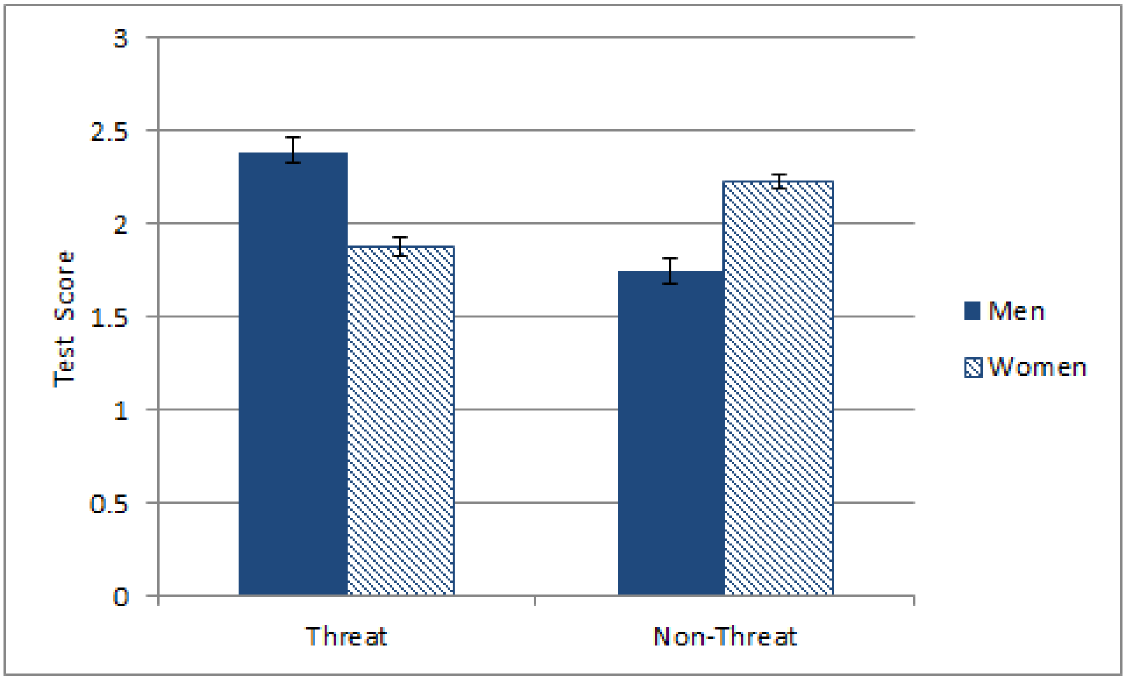
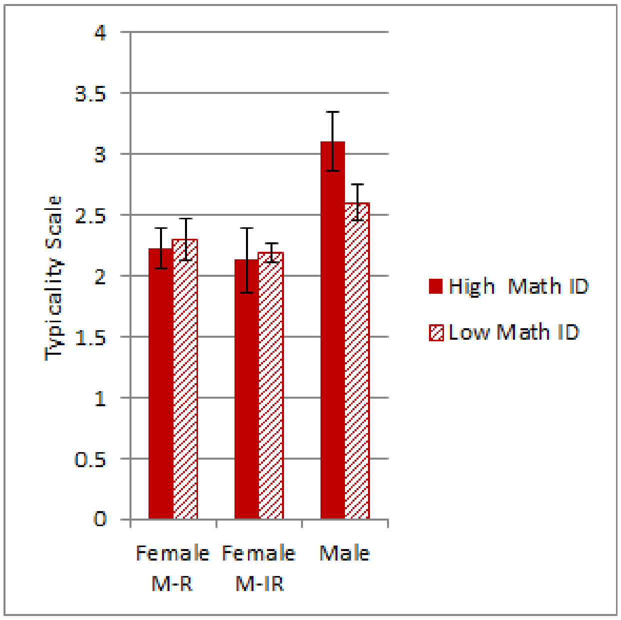
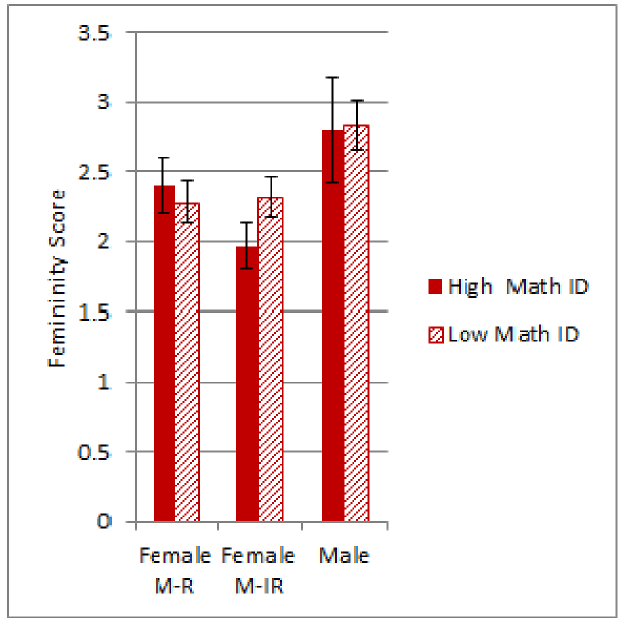


Figure 3. Number correct on math test as a function of gender and stereotype threat condition. The condition * gender interaction was significant (p=.03).

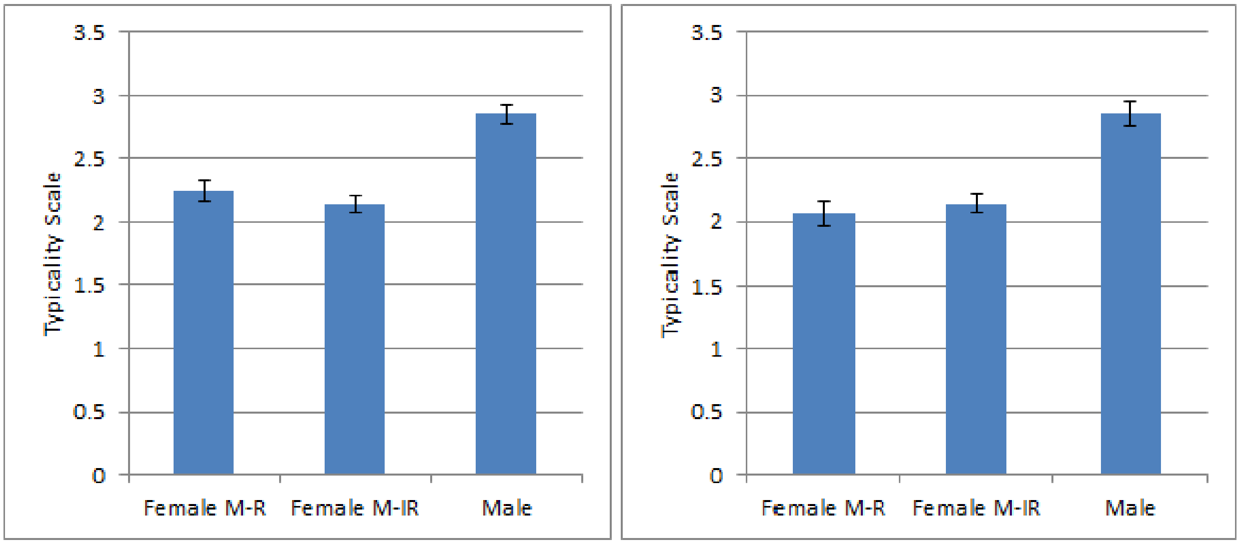


Explicit Analysis



Implicit Analysis

Figure 4. Ratings of trait typicality on a scale of female (low numbers) to male (high numbers) as a function of trait type, condition and either explicit (left graph) or implicit (right graph) math ID. Only threat condition data is depicted here. The three-way interaction was not significant. The difference between high and low math ID ratings of Female M-IR traits was non-significant on both explicit and implicit levels.



Explicit Analysis Implicit Analysis

Figure 5. Significant main effects of trait type on ratings of typicality from female (low numbers) to male (high numbers) resulting from explicit (left graph) or implicit (right graph) analyses. Planned comparisons between male and female traits are significant. Planned comparisons between Female M-R and Female M-IR traits are marginally significant in the explicit analysis, but non-significant in the implicit analysis.

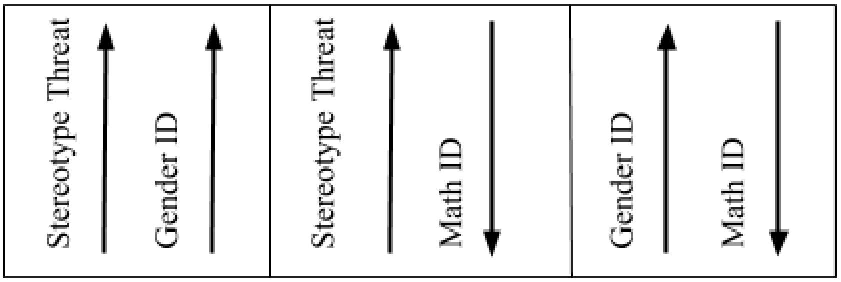


Figure 6. Relationship between gender ID, math ID and mathematics-related stereotype threat in women. Results contrary to hypothesis. Not all relationships depicted here were significant.

Appendix A

Implicit Association Test (IAT) Stimuli List

Category	**Words**
Male	Adam, Harry, Josh, Fred, Mike
Female	Amanda, Sara, Katie, Meredith, Betsy
Math	algebra, geometry, computation, numbers, equations
Language	English, grammar, adjective, words, paragraph
Self	I, me , my, mine, self
Other	they , theirs, them, other, their
Insect	cricket, fly, termite
Flower	lilac, rose, violet
Good	joy, love, pleasant
Bad	cruel, misery, war

Appendix B

IRB Approval

Bard College Institutional Review Board

Date: November 11, 2013
To: Elyse Neubauer
Cc: Kristin Lane
From: Michelle Murray
Re: The Relationship Between Gender Identification and Stereotype Threat

DECISION: APPROVED (with conditions)

Dear Elyse,

The Bard Institutional Review Board reviewed your proposal at our November meeting. Your proposal is approved through November 11, 2014, **with the following condition**:

- Consent Form: please include contact information for your advisor on the consent form.

Once the revised consent form is submitted to mkmurray@bard.edu and irb@bard.edu, and you receive email confirmation from me, you can begin data collection – another letter of approval from the Bard IRB is not necessary.

Please notify the IRB if your methodology changes, or unexpected events arise.

We wish you the best of luck with your research.

Michelle Murray
mkmurray@bard.edu
IRB Chair