5-4-2019

Comparing to ingroup and outgroup members: Do we assimilate, contrast, or neither?

Kathryn Bruchmann
Santa Clara University, kbruchmann@scu.edu

Meghan C. Evans

Recommended Citation

Bruchmann, K., & Evans, M. C. (2019). Comparing to ingroup and outgroup members: Do we assimilate, contrast, or neither? The Journal of Social Psychology, 159(3), 313–327. https://doi.org/10.1080/00224545.2018.1472546

This is an Accepted Manuscript of an article published by Taylor & Francis in The Journal of Social Psychology on May 3, 2019], available online: http://www.tandfonline.com/10.1080/00224545.2018.1472546.

Comparing to ingroup and outgroup members: Do we assimilate, contrast, or neither?

Kathryn Bruchmann
Meghan C. Evans

Santa Clara University

Corresponding Author:
Kathryn Bruchmann
kbruchmann@scu.edu
500 El Camino Real
Santa Clara, CA 95053

Abstract

Previous work studying social comparisons suggests that people are likely to assimilate to ingroup members (e.g., Ledgerwood & Chaiken, 2007), but can also contrast from ingroup members if outgroup members are present (Blanton, Miller, & Dye, 2002). The present research built upon these findings by including a no-comparison control group to test for true contrast and assimilation effects. Across two studies, women primed with a gender-math stereotype received false feedback about their performance on a math task; and in some conditions, they learned of the performance of ostensible male and/or female co-participants. Relative to a no-comparison control, we did not see evidence of ingroup assimilation in either study. However, in both studies, we found that participants were likely to contrast their self-evaluations away from downward targets, regardless of group membership. This suggests that self-enhancement motivations may be stronger than the drive for ingroup assimilation.

Keywords: social comparison, assimilation, contrast, stereotype threat, control groups

People often rely on social comparison information when evaluating their own abilities or characteristics (Festinger, 1954; Suls, Martin, & Wheeler, 2002). Much comparison research simply examines how people respond to comparison information from better off or worse off others. However, there has been very little research examining how people use social comparison information specifically when experiencing stereotype threat; a search of the literature yields only 15 publications that discuss both fields, and even fewer that examine how threat influences the effects of comparison. This gap in the social comparison literature could have important implications for understanding how stereotype threat can affect people in real world settings. For example, if a woman learns about her performance on a math task, the comparison information might have a different effect depending on whether it comes from another woman or a man, especially if the negative stereotype about women's math abilities is emphasized. If certain comparisons improved her feelings about her math ability, this could buffer against future threats; however, if a comparison resulted in feeling even worse about math ability, threat effects could be compounded for future tasks. Additionally, studying the effects of social comparison in a stereotype threat setting can offer insights as to how both processes influence self-evaluations. Research suggests that comparisons with ingroup and outgroup members can yield different outcomes (e.g., Hall & Crisp, 2008), especially when group membership is important for the domain. As such, the present research seeks to understand how people socially compare with ingroup versus outgroup members, specifically in situations when negative stereotypes about the group are made salient.

Comparisons Under Threat

Much research has focused on how people use social comparison information when experiencing ego threat in general; a large body of work points to the idea that when

experiencing threat, people are more likely to seek out downward comparison targets (Friend & Gilbert, 1973; Hakmiller, 1966; Vohs & Heatherton, 2004; Wills, 1981), and to use comparison information to form more favorable self-evaluations (e.g., Beer, Chester, & Hughes, 2013; Wood, Taylor, & Lichtman, 1985). Overall, this work demonstrates that many kinds of threat lead to people using social comparison in self-enhancing ways. However, there is less known about how people compare when they are threatened by their own social identity.

Stereotype Threat. Stereotype threat effects occur when members of a stigmatized group are faced with a stereotype-relevant task (e.g., Steele, 1997); despite efforts to perform well, people experiencing stereotype threat often confirm the stereotype. For example, when women were told that gender differences were expected on a math test (versus that there were no gender differences on the particular test), they performed worse on the task (Spencer, Steele, & Quinn, 1999). Stereotype threat effects have been found for African Americans and Latino students and academics (Brown & Lee, 2005), elderly adults and reaction time (Lambert et al., 2016), and white students and performance on a race-based implicit association test (Frantz, Cuddy, Burnett, Ray, & Hart, 2004). Importantly for this work, there is also ample evidence of stereotype threat affecting women's performance in cognitive and physical domains as widespread as visuospatial assignments (Campbell & Collaer, 2009), math ability tests, (Tomasetto, Alparone, & Cadinu, 2011), specific athletic tasks (Hively & El-Alayli, 2014), and financial decision making (Carr & Steele, 2010). In addition to affecting performance on stereotype relevant tasks, stereotype threat also leads to disengagement from stereotyped tasks or domains (e.g., Steele, 1997), decreased practice or self-handicapping (e.g., Stone, 2002), or decreased identification with the stereotyped group (e.g., Cohen & Garcia, 2008).

Stereotype threat and social comparison. Relatively little work has focused on how stereotype threat and social comparison can intersect. Some research suggests that comparison targets can actually induce stereotype threat. Hall and Crisp (2008) demonstrated that women (who strongly identified with their gender) performed worse on a math test after being primed with men by writing an essay about things men typically do. However, women who were low-identifiers with their gender performed better after being primed with the concept of men. In other words, they contrasted away from their group prime.

Marx, Stapel, and Muller (2005) demonstrated that women experiencing stereotype threat performed better on a math test after comparing to a woman who was described as very good at math and likely to pursue a quantitative graduate degree than after comparing to a woman who was described as bad at math and likely to pursue a career in acting. In other words, comparing to an ingroup member who defied the stereotype helped buffer against the typical stereotype threat effects.

There is also evidence that social comparison can influence self-evaluations when experiencing a social identity threat. Redersdorff and Martinot (2009) studied psychology majors whose majors were framed as either high status (in comparison to philosophy majors) or low status (in comparison to pre-med majors). When their ingroup was described as being low in status (but not high), participants felt better about themselves if an ingroup member outperformed them. In other words, when part of a lower status (i.e. “stigmatized”) group, participants assimilated their self-evaluations to upward ingroup targets. Related research (Mackie, 1984) suggests that when an ethnic group identity is described as low status, children are more likely to report that their ingroup members are better-off, whereas that their outgroup members are worse-off than they are.

In a study that tested stereotype threat more specifically, Blanton, Crocker, and Miller (2000) had African American women complete a difficult "IQ" measure of "natural math ability" with either another African American woman or a white woman (both of whom were confederates). Participants learned that they scored in the 67th percentile and then found out from overhearing the experimenter that their ostensible co-participant had scored either in the 20th percentile or the 99th percentile before rating their performance self-esteem. They argued that participants assimilated to ingroup members and contrasted from outgroup members. More specifically, when they were comparing to another African American woman (i.e., an ingroup member), they felt better about themselves when they were outperformed. However, when they were comparing to a white woman (i.e., an outgroup member), participants felt worse about themselves when they were outperformed than when they had performed better than the co-participant. Overall, evidence from these three studies suggests that stereotype or social identity threat is likely to lead to ingroup assimilation and outgroup contrast.

In a conceptually related study, Blanton, Christie, and Dye (2002) tested how women felt after comparing to both ingroup members (another woman) or outgroup members (a man). In this study, women completed a math and spatial abilities task and then found out that she had earned a score of 67 out of 100 and that her co-participants had earned a 99 and a 35 out of 100. It was manipulated whether it was the ingroup member who was performing better or worse than the participant. They also manipulated stereotype salience; in the low salience condition, participants were told that most college students could solve the math problems regardless of their training, and in the high salience condition, the experimenter suggested that men might perform better on the task. When the stereotype salience was low, participants felt better about their own performance when their ingroup co-participant had performed better (i.e., when both

the participant and the woman co-participant had done better than the man); however, when the stereotype salience was high participants felt better about their own performance when they had outperformed their ingroup co-participant. The authors refer to this as a kind of shifting standards-- when a stereotype about gender and math ability was made salient, the male target might have become irrelevant, so the women only compared (i.e., contrasted) to the other woman's score. In a second study, they demonstrated that stereotype endorsement moderated how social comparison information was used under threat; women who endorsed the stereotype that men are better at math than women were more likely to "shift standards" and evaluate the self more favorable when outperforming another woman, whereas women who did not endorse this stereotype felt better about themselves when they and another women outperformed a man.

While other studies have demonstrated assimilation to ingroup members when experiencing stereotype threat, the results of Blanton et al. (2002) suggest that contrast from ingroup members might be likely under certain situations; particularly situations in which stereotype salience and endorsement are high. So why these differences? Blanton (2001) offers an explanation that gender may be unique in terms of its effects on group-experience, and therefore on social comparison processes; for example, Blanton argues that in a person-to-person interaction, gender is not as distinct of a feature as something like race or ethnicity so it might not even activate a social identity. However, the research described above by Marx and colleagues (2005) and Hall and Crisp (2008) suggests that gender *can* be a "grouping" factor when gender stereotypes are made more salient, and that women are likely to assimilate to ingroup primes in these situations. Additionally, work by Ledgerwood and Chaiken (2007) suggests that people are more likely to assimilate to ingroup primes when the ingroup is broad

versus narrow. As such, it does seem possible that gender would be considered a relevant and distinct grouping variable, especially when a negative gender stereotype is made salient.

The goal of the present work is to test how women compare to ingroup members when outgroup members are absent versus present. Based on work described above, we would predict that when comparing only to ingroup members (i.e., other women), women should assimilate their self-evaluations to their comparison targets; however, when women are comparing to both ingroup and outgroup members, we would predict that women would "shift standards" (as seen in Blanton et al., 2002) and contrast away from their ingroup members. Ledgerwood and Chaiken (2007) demonstrated that primes result in assimilation when they are providing context, but contrast when providing a distinct point of comparison. For gender, we argue that the presence of ingroup members only might provide "context", but the presence of an outgroup member could shift the focus of the ingroup member to become a distinct point of comparison or even competition.

True Contrast and Assimilation? An additional alternative explanation for the findings discussed above is that results were misinterpreted as being assimilation and contrast. Recent work (Bruchmann, 2017) demonstrates that without a no-comparison control group, it is impossible to make claims of true assimilation or contrast. It is possible that one or more of the conditions in these studies would not be different from a baseline self-evaluation. Perhaps, given the tendency to self-enhance when experiencing threat, people would be more likely to contrast away from downward targets than upward and more likely to assimilate to upward targets than downward-- regardless of group membership. Including a no-comparison control condition would further the understanding of how comparison information is used when in situations where stereotypes are activated.

The Present Research

The present study tests how women compare to ingroup and outgroup members both independently and together. After being primed with gender stereotypes, female participants completed a math test and learned they earned a mid-range score (in the control condition, this is all they learned). Some participants also learned that they had performed better or worse than a female or male co-participant-- in other words, they compared only to an ingroup or an outgroup member who had done better or worse. In these conditions, based on the prior work suggesting ingroup assimilation and outgroup contrast, we expected that participants' self-evaluations be more positive when they were outperformed by ingroup members or had outperformed an outgroup member, and less positive when they were outperformed by outgroup members or had outperformed an ingroup member. However, by including no-comparison control group, we will be able to determine whether assimilation and contrast is happening in both directions. In other conditions, participants learned that their performance was directly in between a female and a male co-participants'; that is, participants compared to both ingroup and outgroup members. In these conditions, we expected that participants would have more positive self-evaluations when performing better than their ingroup member than when performing worse than ingroup members. If the presence of an outgroup member makes people more likely to contrast away from ingroup members (the idea of "shifting standards"), then we would predict that people evaluate themselves more positively (than control) when they outperform an ingroup member when an outgroup member is present, and less positively (than control) when an outgroup member is absent.

Experiment 1

Method

Participants and Design. Female undergraduate students (N=163, 42.5% non-white, M_{age}= 18.84 years, SD_{age}=.95 years) participated in exchange for partial course credit. Participants were either at the lab session alone, or were with another female co-participant, a male confederate, or both. Participants were randomly assigned to one of seven conditions where they compared themselves to no one (control), a female who performed better (female upward) or worse (female downward), a male who performed better (male upward) or worse (male downward), both a female who performed better and a male who performed worse (female upward and male downward), or both a male who performed better and a female who performed worse (male upward and female downward).

Materials and Procedure. Participants were informed by a female experimenter that they would be helping to standardize a new type of math test known as modular arithmetic for a college sample (cover story adapted from Van Loo, Boucher, Rydell, & Rydell, 2013). To induce stereotype threat, the experimenter went on to explain that preliminary research has suggested that modular arithmetic problems provided a better assessments of "natural math ability" than other math tests, but that the data typically demonstrate a gender difference "such that men typically perform better than women" and that the present study was aiming to learn more about the reason behind this gender difference (manipulation adapted from Beilock et al., 2007, & Blanton et al., 2000).

Modular Arithmetic Exam. Participants began the study by learning how to complete modular arithmetic (MA). MA equations take the form of $a \equiv b \pmod{c}$ and are solved by subtracting b from a and then dividing by c. If the resulting answer is an integer, the answer is "true", and if it is not an integer, the answer is "false". After learning the process and going

through two practice items, participants completed 50 MA items (adapted from Beilock et al., 2007, and Van Loo et al., 2013).

Social Comparison Feedback. When participants had all completed the MA exam, the experimenter told them that she would be retrieving their score reports from a printer in the next room; upon returning, she passed out a score report to each present participant with their own score highlighted. In reality, the score-reports were all pre-printed and a bogus score of 81% correct was highlighted. In the control condition, this is the only information that was indicated on the score report. In the experimental conditions, participants could also see the (bogus) scores and gender of any other participants. The co-participants' scores were either upward comparisons (98%), downward comparisons (64%), or both upward and downward comparisons in the conditions where two others were present. See Table 1 for complete score report information across conditions for both studies.

***Dependent Measures*.** Next, on pencil and paper, participants completed the dependent measures. First, they indicated how they would rate their performance, and how satisfied they were with their performance on the MA task (1= *extremely poor/dissatisfied*, 7= *extremely good/satisfied*). Participants also indicated how they would rate their ability on modular arithmetic, and on math in general (1= *extremely poor*, 7= *extremely good*).

State Self-Esteem. Participants next completed the 7-item performance subscale from the State Self-Esteem Scale (Heatherton & Polivy, 1991); participants indicated their agreement with items including "I feel confident about my abilities" and "I feel as smart as others" (1= *not at all*, 5= *extremely*).

Perceptions of Modular Arithmetic. Next, participants rated how much they liked modular arithmetic as a way to measure math ability (1= *dislike very much*, 5= *like very much*),

and how much they liked modular arithmetic compared to other math problems (1= *strongly prefer modular arithmetic*, 5= *strongly prefer standard math*). Finally, participants were probed for suspicion, and were debriefed.

Results

Modular Arithmetic Performance. The modular arithmetic items were scored and accuracy was calculated dividing the total score by 50 (the total number of items). On average, participants answered 85.51% (*SD*=15%) of the items correctly (significantly higher than the bogus feedback of 81%, $t(163)= 3.95$, $p<.001$). Accuracy was marginally different across comparison conditions, $F(6, 156)= 1.86$, $p=.092$, $\eta_p^2=.067$, however posthoc comparisons (with Bonferroni corrections) did not reveal differences between conditions. We included MA accuracy as a covariate in all other analyses and the patterns were not different, so differences in accuracy will not be discussed further.

Self Evaluations. See Table 2 for means and standard deviations for Study 1 dependent measures. We first calculated a score performance self-esteem subscale from the SSES (Heatherton & Polivy, 1991; $\alpha=.79$) and conducted an ANOVA. No significant effect of condition emerged, $F(6, 154)= 1.42$, $p=.211$, $\eta_p^2=.052$.

As a secondary test, we averaged participants' ratings of their performance, satisfaction, and MA ability to create a self-evaluation composite ($\alpha= .76$). An ANOVA revealed a main effect of comparison condition on self-evaluations, $F(6, 155)= 3.75$, $p=.002$, $\eta_p^2= .127$. Planned comparisons were conducted to test for the specific *a priori* predictions described above.

Did women assimilate to other women? No, we did not see evidence of assimilation. Participants who compared to an ingroup upward target ($M=3.94$, $SD= .85$) rated themselves less favorably than the baseline control ($M=4.49$, $SD= .76$; $p=.011$, $d= -.68$). Additionally, when

compared to the no-comparison control, participants rated themselves marginally more favorably when they had an ingroup downward target (*M*=4.86, *SD*= .55,*p*=.086, d= .56). In other words, instead of finding assimilation to ingroup members like predicted, we saw distinct patterns of contrast from ingroup members. Participants rated themselves as being different from co-participants who were also women. Even if we did not include a baseline control group, we would not have seen what looked like patterns of assimilation: participants comparing to only a female co-participant provided less favorable self-evaluations when comparing to an upward comparison target than when comparing to a downward comparison target (*p*<.001, *d*= -1.29).

Did women contrast from men? No. Participants who compared only to male confederates did not rate themselves differently from the control (*p*s=.349 & .107, *d*s= -.64 & .30) whether it was downward comparison information (*M*=4.87, *SD*= .63) or upward comparison information (*M*=4.28, *SD*= .84). In other words, finding out that they had performed better or worse than an outgroup member did not yield different self-evaluations than finding out no social comparison information at all. Had we not included our control group, we would have observed a pattern that might have looked like contrast; participants comparing to just a male confederate rated themselves more favorably when they had downward comparison information than upward *p*=.023, *d*= .79).

Did women "shift standards" when they compared to women and men? No. Participants rated themselves similarly when they had both ingroup and outgroup comparisons whether it was the female co-participant who had performed better (*M*=4.41, *SD*= .98) or the male confederate who had performed better (*M*= 4.47, *SD*= .75, *p*= .805, *d*= -.07). Additionally, participants did not rate themselves differently from control when they had both ingroup and outgroup comparisons, regardless of the comparison direction (*p*'s= .699 & .909, *d*s= .10 & .03).

Does comparison information from a man change how comparison information from a woman is used? Maybe, but not in the predicted direction. The bidirectional comparison conditions did appear to dampen any effects of the unidirectional comparison conditions. When participants compared to both an upward female target and downward male target, they rated themselves more favorably than when they compared only to an upward female target (p=.038, d= .51). In other words, outperforming an outgroup member seemed to buffer against the effects of underperforming compared to an ingroup member. When participants compared to both an upward male target and a downward female target, they rated themselves marginally less favorably (p=.092, d= -.59) than when comparing just to a downward female target; in this case, underperforming an outgroup member seemed to take away the benefit of outperforming an ingroup member.

Perceptions of Modular Arithmetic. Participants' general rating of MA and their rating of MA compared to standard math were averaged to create a composite (α=.71) in which higher scores indicate a more favorable perception of MA. A significant main effect of comparison condition emerged, $F(6, 155)= 2.65$, p=.018, η_p^2=.093. Post hoc analyses (using Bonferroni corrections) suggested that participants who had outperformed a female co-participant (M=3.19, SD= 1.02) had more positive perceptions of the modular arithmetic task than participants who had been outperformed by a male confederate (M=2.16, SD= 0.82, p=.011, d= 1.11). Significant differences between the other conditions did not emerge.

Discussion

Overall, our results did not support our predictions that participants would assimilate to ingroup members unless an outgroup member was present. Instead, we saw evidence of contrast effects; specifically, participants felt marginally better when they outperformed an ingroup

member, and felt worse about their own performance (versus the control) when outperformed by an ingroup member. This finding is inconsistent with previous research that has demonstrated assimilation to ingroup members (Blanton et al., 2000; Hall & Crisp, 2008; Ledgerwood & Chaiken, 2007; Redersdorf & Martinot, 2009), and is inconsistent with a pattern of self-enhancement in which people might be more likely to assimilate to upward targets and contrast from downward. One potential explanation for our different results was our population. For example, Blanton et al. (2000) studied African American women who were comparing to other African American women or white women when they found patterns of ingroup assimilation. As suggested by Blanton (2001), race might function differently from gender as a cue of similarity that could lead to assimilation or contrast. However, other research suggests that gender should be experienced in the same way as any other grouping variable (e.g., Hall & Crisp, 2008; Marx et al., 2005; Ledgerwood & Chaiken, 2007).

Additionally, we saw only weak evidence that people contrast from outgroup members. Participants rated themselves more favorably when they outperformed a male co-participant (versus control), but did not rate themselves less favorably (versus control) when outperformed by a male co-participant. This, again, is inconsistent with previous research that suggests we are more likely to contrast away from outgroup members (Blanton et al., 2000; Hall & Crisp, 2008; Ledgerwood & Chaiken, 2007; Redersdorf & Martinot, 2009). Although, this pattern is consistent with self-enhancement in that participants did not report feeling worse than a baseline control when comparing with better-or-worse-off outgroup members. One potential explanation for our pattern of results was that we used a no-comparison control group. If the present study had not included this control group, we would see patterns that seem to resemble contrast effects when comparing to outgroup members; performing better than a man led to more favorable self-

evaluations than performing worse than a man. However, the inclusion of the no-comparison control group demonstrates that only comparing to a worse-off target led to contrast effects, and that comparing to a better off target did not negatively affect self-evaluations. Given this difference between our study and the others described, it is possible that the original studies were misinterpreted.

Finally, we did not see evidence that comparing to both ingroup and outgroup members changed the way ingroup comparison information was used. We hypothesized that the presence of an outgroup member would lead people to contrast away from their ingroup members, whether through a process of competition or because they provided a more relevant point of comparison. However, self-evaluations in the conditions where both ingroup and outgroup members were not different from each other or baseline control group. Beyond this, when comparing the bidirectional and unidirectional conditions, we saw evidence that the outgroup member was the comparison target that mattered more than the ingroup member and not the other way around; for example, performing better than a male co-participant and worse than a female co-participant led to more favorable self-evaluations than just performing worse than a female co-participant. This pattern of results is the opposite of what we would have predicted based on prior research by Blanton et al. (2002) that suggested that when stereotype salience or endorsement was high, women "shifted standards" to contrast with other women when comparison information with men was also present.

One potential issue with our study was that we did not measure whether our participants were aware of or endorsed the stereotype that women are bad at math. This is an issue because if our participants were not aware of the stereotype, our stereotype threat manipulation might not have been effective, and gender might not have been a salient enough feature to lead to

assimilation effects. Additionally, Blanton et al. (2002) demonstrated that shifting standards was more likely for women who endorsed this stereotype. Considering that women who were told they were outperformed by men had more negative attitudes about the modular arithmetic task than women who thought they had outperformed another woman, it is possible that in general our participants did not endorse the stereotype that men are better at math.

Study 2

Overview

The purpose of Study 2 was to provide an additional test of whether women would assimilate to ingroup comparison information and contrast from outgroup situation when a gender stereotype was made salient. In Study 2, we adjusted our comparison information so that it was given in the form of percentiles and not just raw scores. The use of university-wide percentile feedback information rather than percentage score feedback from a present co-participant allowed for a standardization of all conditions so that only one participant was involved in the experiment at a time. Additionally, we hoped that the percentile information would make the comparison feedback more ambiguous which should allow for assimilation.

Additionally, to make stereotype threat more salient, all experimenters in Study 2 were male. Keller and Sekequaptewa (2008) demonstrated that threat can be induced for women if they are the only women in the room; as such, having a male experimenter should have induced threat for all participants, whereas in Study 1, participants were never the only woman in the room. We also altered several items on the MA task from Study 1 that were answered correctly by most participants in order to make them more difficult, and in order to make the task more threatening. Finally, we included measures of ingroup identification and stereotype endorsement as potential moderators of the social comparison effects.

Method

Participants and Design. Female undergraduate students (N=86, 52.2% non-white, M_{age}= 18.78 years, SD_{age}= 0.94 years) participated in exchange for partial course credit. Participants arrived at the session alone, were greeted by a male experimenter and were randomly assigned to one of five comparison conditions: control, female upward, female downward, male upward, male downward.

Materials and Procedure. Study 2 followed the same procedures as Study 1 except for the following changes.

Modular Arithmetic Exam. Participants learned how to complete MA items through the same instructions as Study 1. The only difference between the MA exam for Study 2 was that 12 of the 50 items were altered to increase their difficulty.

Social Comparison Feedback. When participants had completed the MA exam on the computer, they read that they would be shown percentile infographics displaying how their performance compared to their university's female and male student populations; they read that they should pay careful attention to this information because they would be asked questions about it.

On the next screen, the infographic was pre-generated with bogus percentile feedback information. Participants always learned that they scored 81% correct on the MA task (as in Study 1); depending on condition, they also learned that they were in the 53rd or 91st percentile for males or females at their university. While two boxes always appeared as if two infographics (one for male and one for female) would be displayed, one of the boxes would be grayed out as if the image could not load on the computer (See Figure 1); in other words, participants received only male or female percentile feedback information to compare to, never both. In the control

condition, it appeared as if neither image could load and they received no comparison information. If the participant brought the image "issue" to the attention of the experimenter, the experimenter responded as if it was an accident and say, "Oh that's strange. You are supposed to get information about the breakdown for both males and females. I guess just keep going and I'll make sure the principal researcher knows about this."

Dependent Measures.

Self-evaluations. Following the infographic display, participants completed self-evaluations on the computer. As in Study 1, they rated their MA performance, satisfaction with their performance on the MA task, and general MA ability (1= *extremely poor/dissatisfied*, 7= *extremely good/satisfied*). Furthermore, they rated their general math ability in comparison to female and male students at their university, and in general (1= *much worse*, 7= *much better*). Participants also completed the 7-item performance self-esteem measure from Study 1 (Heatherton & Polivy, 1991).

Stereotype endorsement. Participants indicated whether they were aware of any stereotypes about gender and math (*yes* or *no)*, and how much they agree with the statement that "women are worse at math than men" (1= *completely disagree*, 7 = *completely agree).* Participants also indicated how important their gender and math ability were to their identity (1= *extremely unimportant*, 7= *extremely important*). Finally, participants were asked to recall their score and percentile information, were probed for suspicion, and were debriefed.

Results & Discussion

Manipulation Check. All but six participants accurately recalled their score on the modular arithmetic task. The six participants who mis-remembered instead recalled their

percentile information and indicated that they scored either 53% or 91%. Removing these participants from the data set did not change the results of the study, so they were left in.

MA Accuracy. Despite attempts to make the modular arithmetic task more difficult in the second study, participants actually performed better on it than they did in the first study. On average, participants answered 88.18% of the items correctly (*SD*= 11.81%); accuracy did not differ across conditions, $F<1$.

Self Evaluations. As in Study 1, we calculated a score on the performance self-esteem measure (α= .81); however, there was no evidence of an effect of the comparison manipulation on state self esteem, $F<1$. Again, we averaged participants' ratings of their performance, satisfaction, and MA ability to create a self-evaluation composite (α= .87). An ANOVA revealed a significant effect of comparison condition on self-evaluations, $F(4, 81)= 10.66$, $p<.001$, $\eta_p^2= .345$. Specific planned comparisons were conducted based on our a priori research questions. See Table 3.

Did women assimilate to other women? No, as in Study 1, we did not see evidence of assimilation. Participants with downward female comparison information (*M*=5.26, *SD*= 1.32) rated themselves more favorably than the control (*M*= 3.80, *SD*= 1.09, $p<.001$, $d= 1.21$) but participants with upward female comparison information (*M*=4.20, *SD*= 1.01) did not rate themselves differently from the control group ($p=.291$, $d= .38$). In other words, we do not see symmetrical contrast effects; instead, we see that people are more likely to contrast away from downward comparison information than upward.

Did women contrast from men? Yes, but only when they outperformed their outgroup members. Participants with only male percentile information rated themselves more favorably when they had downward comparison information (*M*= 5.73, *SD*= .77) than if they did not have

comparison information (p<.001, d= 2.05). Participants who had male upward comparison information (M= 3.98, SD= 1.10) did not rate themselves differently from control (p=.634, d=.16). As with the participants with female percentile information, only the downward comparison condition was different from control.

Whether participants were comparing with the ingroup or the outgroup, we see patterns of self-enhancement. Overall, relative to control, participants rated themselves no differently after comparing to upward targets, but more favorably after comparing to downward targets.

Did stereotype awareness or endorsement moderate these effects? No. While 73.3% of participants indicated that they were aware of a stereotype about gender and math, there was not a significant stereotype awareness x comparison condition interaction, $F(4, 75)=1.05$, p=.390, η_p^2=.053. Additionally, stereotype endorsement did not moderate the effects on self-evaluations, $F(15,62)= 1.48$, p=.14, η_p^2= .264. However, this could be due to a floor effect; 41.9% of participants indicated that they "completely disagreed" with the stereotype that women are worse at math than men. Only 7% of participants indicated any endorsement ("slightly agree") of the stereotype.

As a secondary test, we examined whether being a STEM major moderated the comparison effects on self-evaluations; there is evidence that women who are more strongly identified with math experience stronger threat effects (e.g., Nguyen & Ryan, 2008). Again, there was no evidence of either a main effect of being a STEM major, or a STEM major x comparison condition interaction, Fs<1.

Relative Math Ability. We did not find evidence of an effect of comparison condition on participants' ratings of their math ability compared to their typical female classmate, or to the typical man or woman in general (Fs<1). However, an ANOVA did reveal a main effect of

comparison condition on participants' ratings of their math ability compared to their typical male classmate, $F(4, 79)= 2.52$, $p=.048$, $\eta_p^2= .113$. Post hoc tests (using Bonferroni corrections) indicate that participants who received downward male percentile information were marginally more likely to agree with the statement that they had better math ability than their male classmates ($M= 4.71$, $SD= .92$) than participants who received female upward percentile information (M=3.76, SD= 0.97; $p=.068$, $d=1.00$). This response is logical considering that these participants received feedback telling them they had better ability than their male classmates, however, we do not see similar "logical" effects for students in the other conditions.

Better than Average Effects. One potential reason that our comparison conditions generally did not affect people's ratings of their math abilities is that people are instead demonstrating better than average effects. Responses to the comparative questions were subjected to a one-sample t-test compared to the value of 4 (the midpoint of the scale). Participants across conditions rated themselves as having better math ability than the average woman at their university ($M= 4.24$, $SD= 1.02$, $t(84)= 2.13$, $p=.036$, $d=.24$), and better than the average woman in general ($M= 4.51$, $SD= 1.06$, $t(85)= 4.48$, $p<.001$, $d=.49$). However, participants did not rate their math ability differently from the average man at their university ($M= 4.04$, $SD= 1.02$, $t(83)= .32$, $p=.750$, $d=.04$) or from the average man in general ($M= 4.18$, $SD= 1.03$, $t(84)= .84$, $p=.116$, $d=.18$).

While most participants did not explicitly report endorsing the gender math stereotype, these comparative judgments tell a different story; participants think they have equal ability to most men, but better ability than most women suggesting that they do think there is a gender difference in math ability. As such, we tested the relative math ability judgments as a moderator for the social comparison effects, but no significant effects emerged ($F<1$).

General Discussion

Across two studies, we tested how women use social comparison information from other women and/or from men after completing a threatening math test. Overall, we did not see strong evidence that the group membership of the comparison targets affected self-evaluations. Nor did we see evidence that people used comparison information from ingroup versus outgroup members differently. Instead, we saw evidence that comparing to worse-off others-- whether they were ingroup or outgroup members-- led to more favorable self-evaluations.

Our results our inconsistent with previous research that suggests that people are likely to assimilate to similar or ingroup comparison targets (e.g., Blanton et al., 2000; Mussweiler et al., 2004). Additionally, our results were inconsistent with prior work that demonstrates that women contrast from other women in stereotype threat situations when men are also present (Blanton et al., 2002). Our studies, and our different results, could be due to a number of factors. First, across both studies, we were relying on gender as our grouping variable. Most of the work on intergroup comparisons described above does not examine gender, and some argue that gender is not a distinct enough social category to indicate group membership (Blanton, 2001). As such, future research should test this idea more specifically by comparing gender groupings to other types of groupings (e.g., by race or ethnicity).

Second, our studies were different specifically from research examining stereotype threat and social comparison in that we did not manipulate stereotype threat; instead, all of our participants were in a situation designed to induce stereotype threat. In both studies, we used threat paradigms and materials that have been successful for other researchers, but it is possible that our participants did not experience stereotype threat as we designed. Indeed, most of our participants did not report endorsing the stereotype that women are not as good at math as men

(Study 2), and were less likely to believe our math task to be credible if they were outperformed by a man (Study 1). According to Nguyen and Ryan (2008), subtle manipulations of threat work better for the gender/math stereotype, so perhaps the present manipulation (though adapted from other research) was too overt for this population. It is possible that gender could be a grouping variable if it is made salient through a stereotype threat manipulation, so future research should examine this as a potential explanation for why our results were inconsistent with previous research.

Third, when considering how our findings compare to those of Blanton et al. (2002), it may be relevant to consider the potential shift in gender stereotype awareness over the last 15 years. It is possible that the social context surrounding these gender stereotypes has changed reflecting social progress. While findings from Blanton et al. (2002) suggest that when women reject the stereotype that men are better at math, they are more likely to assimilate to other women and thereby feel a boost of pride in their social identity, in 2018 those assimilation effects are not so evident. However, this may reflect that women no longer use gender as a frame of reference when making comparisons about their own abilities in various tasks. Indeed, we did not see evidence that participants explicitly endorsed the stereotype that men are better at math than women. However, we did still see some evidence that this stereotype is endorsed at a more implicit level; our participants considered themselves to be better than other women at math, but not any different in skill level than men. Indirectly, this suggests that they believe at some level that men are more skilled at math than women. It will be important for future research to consider how stereotypes surrounding gender and math continue to change over time.

Self-Enhancing Social Comparisons

We did see evidence consistent with prior research that people are likely to contrast away from their outgroup members. And, more specifically, through the use of a no-comparison control group, we mostly saw evidence of asymmetrical contrast effects; that is, participants were more likely to contrast away from downward comparison targets than upward targets. People often argue that having an upward comparison as a role model is a strong intervention for stereotype threat (e.g., Herrmann, Adelman, Bodford, Graudejus, Okun, & Kwan, 2016); however, the present research would suggest that having downward comparisons (regardless of gender) might actually make women feel better about their math abilities. Across both studies, we found results that are consistent with self-enhancement. While it is not clear that our participants were experiencing threat, there is a plethora of research suggesting that self-enhancement is a common outcome of social comparisons even in the absence of a threat; people are particularly likely to use comparison information from downward comparison targets to self-enhance or repair self-esteem (e.g., Wills, 1981). These kinds of self-enhancing downward comparisons can happen even with irrelevant comparison targets; Zell, Alicke, and Strickhouser (2015) found that people evaluate themselves more favorably after comparing to a downward comparison target even if that target is considered incompetent. Downward targets in Study 1 were scoring 64%, not a performance that would be considered "good" by participants, yet we still see that participants rate their own abilities more favorably after comparing to the downward targets than a baseline control.

We also see evidence of self-enhancement in the conditions in which participants received upward comparison information. Barring one exception, participants in both studies who received upward comparison information did not feel any worse about themselves than participants in the no-comparison control group. This type of self-enhancement would not be

visible without the no-comparison control group, again highlighting how important control groups are for interpreting social comparison effects (see Bruchmann, 2017; Gerber, Wheeler, & Suls, 2018). Yet, self-enhancing after comparing to upward targets is not uncommon. When faced with threatening upward social comparison information, people often self-repair or self-enhance by improving future performance (Johnson, 2012), discounting the upward comparison target (Alicke, LoSchiavo, Zerbst, & Zhang, 1997; Lockwood & Kunda, 1997), or disengaging from the task (see Johnson, 2012 for a review). Additionally, in Study 2, we saw evidence of self-enhancement through Better than Average effects; our participants indicated that they were better at math than most women in general and at their university, and had the same level of math skills as most men in general and at their university.

Conclusion

Across these two studies which built upon previous social comparison and stereotype threat research, we did not see evidence for assimilation to ingroup members or a shifting of standards driving contrast from ingroup members. Instead, we saw strong and consistent evidence of contrast away from comparison targets regardless of group membership; additionally, because of our control group, we saw evidence of mostly downward contrast, not upward. In other words, comparing to worse-off others led to more favorable self-impressions than a baseline control, whereas comparing to better-off others did not generally lead to less favorable self-impressions. While previous research would suggest that the group membership of the comparison targets matters for self-evaluations, our studies provide evidence that it is the direction of the comparison targets that matters for self-evaluations.

References

Alicke, M.D., LoSchiavo, F.M., Zerbst, J., & Zhang, S. (1997). The person who outperforms me

is a genius: Maintaining perceived competence in upward social comparison. *Journal of Personality and Social Psychology, 73*(*4*), 781-789.

Beer, J.S., Chester, D.S., & Hughes, B.L. (2013). Social threat and cognitive load magnify self-enhancement and attenuate self-deprecation. *Journal of Experimental Social Psychology, 49,* 706-711

Beilock, S.L., Rydell, R.J., & McConnell, A.R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill-over. *Journal of Experimental Psychology. General, 133*, 256-276.

Blanton, H. (2001). Evaluating the self in the context of another: The three-selves model of social comparison assimilation and contrast. In G. B. Moskowitz (Ed.), *Cognitive Social Psychology: The Princeton Symposium on the Legacy and the Future of Social Cognition* (pp. 75-87). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Blanton, H., Christie, C., & Dye, M. (2002). Social identity versus reference frame comparisons: The moderating role of stereotype endorsement. *Journal of Experimental Social Psychology*, *38*, 253-267

Blanton, H., Crocker, J., & Miller, D.T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *Journal of Experimental Social Psychology*, *36*, 519-530.

Brown, R. P. & Lee, M.N. (2005). Stigma consciousness and the race gap in college academic achievement. *Self and Identity, 4(2),* 149-157.

Bruchmann, K. (2017). Compared to what? The importance of control groups in social comparison research. *Basic and Applied Social Psychology, 39(2),* 91-100.

Campbell, S.M., & Collaer, M.L. (2009). Stereotype threat and gender differences in

performance on a novel visuospatial task. *Psychology of Women Quarterly, 33*(4), 437-444.

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

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

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(*2*), 117-140.

Frantz, C.M., Cuddy, A.C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30(12),* 1611-1624.

Friend, R.M. & Gilbert, J. (1973). Threat and fear of negative evaluation as determinants of locus of social comparison. *Journal of Personality, 41*(*2*), 328-340

Gerber, J.P., Wheeler, L., & Suls, J., (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, *144*(2), 177-197.

Hakmiller, R. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology*, *1,* 32-39

Hall, N.R., & Crisp, R.J. (2008). Assimilation and contrast to group primes: The moderating role of ingroup identification. *Journal of Experimental Social Psychology*, *44*, 344-353.

Heatherton, T.F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60(6),* 895-910.

Herrmann, S.D., Adelman, R.M., Bodford, J.E., Graudejus, O., Okun, M.A., & Kwan, V.S.Y. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Journal of Basic and Applied Psychology, 38(5)*, 258-268.

Hively, K., & El-Alayli, A. (2014). 'You throw like a girl:' the effect of stereotype threat on women's athletic performance and gender stereotypes. *Psychology of Sport and Exercise, 15*(1), 48-55.

Johnson, C. (2012). Behavioral responses to threatening social comparisons: From dastardly deeds to rising above. *Social and Personality Psychology Compass, 6*(*7*), 515-524.

Keller, J. & Sekaquaptewa, D. (2008). Solo status and women's spatial test performance: The role of individuation tendencies. *European Journal of Social Psychology, 38,* 1044-1053.

Lambert, A.E., Watson, J.M., Stefanucci, J.K., Ward, N., Bakdash, J.Z., & Strayer, D.L. (2016). Stereotype threat impairs older adult driving. *Applied Cognitive Psychology, 30(1),* 22-28.

Ledgerwood, A., & Chaiken, S., (2007). Priming us and them: Automatic assimilation and contrast in group attitudes. *Journal of Personality and Social Psychology, 93*(*6*), 940-956.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73(1),* 91-103.

Mackie, D. (1984). Social comparison in high- and low-status groups. *Journal of Cross-Cultural Psychology, 15(4),* 379-398.

Marx, D.M., Stapel, D.A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology*, *88*(*3*), 432-446.

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology, 87(6),* 832-844.

Nguyen, H.D., & Ryan, A.M. (2008). Does stereotype threat affect test performance of minorities and women? A meta analysis of experimental evidence. *Journal of Applied Psychology, 93*(*6*), 1314-1384.

Redersdorff, S., & Martinot, D. (2009). Being outperformed in an intergroup context: The relationship between group status and self-protective strategies. *British Journal of Social Psychology, 48,* 275- 294.

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. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52(6),* 613-629.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin, 28*(*12*), 1667- 1678.

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, *11*(*5*), 159-163.

Tomasetto, C., Alparone, F.R., & Cadinu, M. (2011). Girl's math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47*(4), 943-949.

Van Loo, K.J., Boucher, K.L., Rydell, R.J., & Rydell, M.T. (2013). Competition in stereotyped domains: Competition, social comparison, and stereotype threat. *European Journal of Social Psychology, 43*, 648-660.

Vohs, K.D., & Heatherton, T.F. (2004). Ego threat elicits different social comparison processes

among high and low self-esteem people: Implications for interpersonal perceptions. *Social Cognition*, *22*(*1*), 168-191.

Wills, T.A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, *90*(*20*), 245-271.

Wood, J.V., Taylor, S.E., & Lichtman, R.R. (1985). Social comparison in adjustment to breast cancer. *Journal of Personality and Social Psychology, 49*, 1169-1183.

Zell, E., Alicke, M.D., Strickhouser, J.E. (2015). Referent status neglect: Winners evaluate themselves favorably even when the competitor is incompetent. *Journal of Experimental Social Psychology, 56*, 18-23.

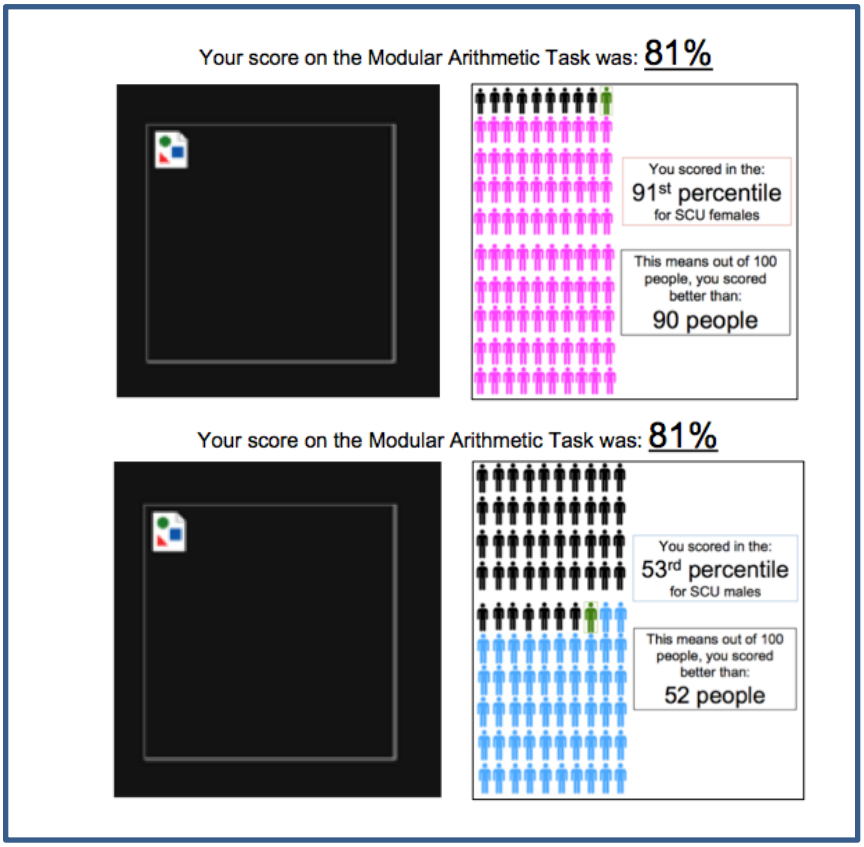

Figure 1. Example comparison feedback for Study 2; female-downward and male-upward conditions.

Table 1

Score and Comparison Feedback across Study 1 and Study 2

	Study 1			Study 2		
Condition	Self	Female Comparison	Male Comparison	Self	Female Comparison	Male Comparison
Control	81%	--	--	81%	--	--
Female Upward	81%	98%	--	81%	91st percentile	--
Female Downward	81%	64%	--	81%	53rd percentile	--
Male Upward	81%	--	98%	81%	--	91st percentile
Male Downward	81%	--	64%	81%	--	53rd percentile
Female Upward, Male Downward	81%	98%	64%	--	--	--
Male Upward, Female Downward	81%	64%	98%	--	--	--

Table 2

Study 1 Means and Standard Deviations across Comparison Conditions

Comparison Condition	Performance Self-Esteem	Self-Evaluation Composite	Perceptions of MA
Control	27.35 (3.32)	4.49 (0.76)	2.90 (0.80)
Female Upward	25.63 (3.51)	3.94 (0.85)	2.91 (0.83)
Male Upward	25.79 (4.08)	4.28 (0.84)	2.16 (0.82)
Female Downward	26.38 (3.37)	4.86 (0.55)	3.19 (1.02)
Male Downward	26.44 (3.29)	4.87 (0.63)	2.58 (1.07)
Female Upward, Male Downward	24.48 (5.23)	4.41 (0.98)	2.54 (1.12)
Female Downward, Male Upward	26.41 (3.81)	4.47 (0.75)	2.80 (0.95)

Table 3

Study 2 Means and Standard Deviations across Comparison Conditions

Comparison Condition	Performance Self-Esteem	Self-Evaluation Composite
Control	24.65 (4.78)	3.80 (1.09)
Female Upward	24.65 (4.31)	4.20 (1.01)
Male Upward	26.18 (4.20)	3.98 (1.10)
Female Downward	26.06 (5.45)	5.26 (1.32)
Male Downward	25.76 (4.55)	5.73 (.77)