UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Acting Woman:
Math Performance and Gender Identity

Joanna Allestam

Supervisor: Eva Ranehill
Master's thesis in Economics, 30 hec
Spring 2023
Graduate School, School of Business, Economics and Law, University of Gothenburg, Sweden

Acknowledgements

I want to thank my supervisor Eva Ranehill who has provided me with guidance, feedback, and encouragement throughout the whole process, from the first research proposal to the finalized version of the thesis. I am truly grateful for your valuable support. I also want to thank my husband, family, and friends for their continuous encouragement.

Abstract

In this paper I causally explore whether perceived math ability has an impact on perceived gender identity. I base my experiment on a framework proposed by Akerlof and Kranton (2000), and randomly assign individuals to receive a signal of relatively high or relatively low math ability. Because math is associated with male stereotypes, I hypothesize that individuals who perceive themselves as good at math identify as more masculine and believe they are perceived as more masculine by others. If true, this indicates women may suffer a gender identity cost for being good at math, which may be contributing to persistent gender gaps in STEM-fields. Including all respondents in my analysis I find no significant effect of treatment on perceived gender identity. When only including participants who consider math to be masculine or gender neutral there is a significant treatment effect indicating those who received a signal of high relative math performance identified as more masculine, and believed they were perceived as more masculine by others. This suggests women who perceive themselves as good at math may experience a gender identity cost, while men may experience a gender identity gain. However, the results are not robust to the inclusion of collected control variables and therefore needs to be confirmed in future research.

Table of Contents

List of Tables

List of Figures

1. Introduction

In 2021 the average gender wage gap amongst the OECD countries was 11.93% (OECD, 2023). In Sweden the gap is narrower, but still present, and was measured to 7.25% the same year (ibid.). It has been found that a reason the gender wage gap still exists is because women are underrepresented in math intensive professions and industries, such as the STEM (Science, Technology, Engineering, and Mathematics) fields (e.g., Blau and Kahn, 2017; Ceci, 2018). To narrow the wage gap, it is therefore important to encourage more women to choose math intensive academic tracks. This argument is strengthened by Joensen and Nielsen (2016) who find that women who take advanced mathematics experience a positive future earnings effect. In 2021, women comprised 63.4% of all university graduates in Sweden (UKÄ, n.d.). Yet in more math intensive academic tracks the share of women was much lower. Amongst the 2021/2022 Master of Science in Engineering[1] graduates only 35% were women (UKÄ and SCB, 2023).

In this paper, I causally explore whether high perceived math ability impacts perceived gender identity by conducting an experiment with Swedish university students. Previous literature suggests that math is associated with a male stereotype (e.g., Cvencek et al., 2011; Nosek et al., 2009; Carlana, 2019). Based on a framework presented by Akerlof and Kranton (2000), I hypothesize that if math is prescribed as a male subject, individuals who perceive themselves as relatively good at math identify as more masculine. They may also believe they are perceived as more masculine by others. If true, a woman who perceives herself as good at math may experience a gender identity cost (i.e., identify as less feminine), while a man may experience a gender identity gain.

In the first part of the experiment, the participants were asked to solve 10 multiple choice math problems. They then received performance feedback indicating their relative performance compared to either a high- or low performing reference group. Being compared to the low performing reference group meant the participants received feedback indicating high relative math performance, and vice versa. Which group they got compared to was randomized with equal probability. After receiving the performance feedback, participants were asked to indicate where they would place themselves, and where others would place them, on a scale

[1] Master of Science in Engineering is used as a translation of the Swedish term "Civilingenjör".

ranging from “Very masculine” = 1 to “Very feminine” = 11. To uncover the participants potential gender stereotypes surrounding math, they were also asked to place the subject mathematics on the gender scale. This revealed that more people in my sample viewed math as masculine compared to feminine, which is in line with previous literature.

Without controlling for the participant’s stereotypes surrounding math, the results show no statistically significant effect of the treatment signaling high relative math performance on gender identity. The results are insignificant for the full sample and for men and women separately. However, when excluding those who indicated they considered math feminine from the analysis, there is a statistically significant effect of treatment on gender identity. This means that, amongst those who consider math to be either masculine or gender neutral, a signal indicating high relative math performance is associated with a shift towards masculine on the gender identity scale. Including an interaction between gender and treatment does not indicate a difference in reaction to treatment between men and women, but this analysis also has less power. Because randomization to treatment was not stratified by gender, women ended up being slightly underrepresented in the group randomized to get a signal of high math performance. The distribution of academic programs also slightly differed between the groups. Therefore, I run my analysis again controlling for gender and academic field. When doing so the previously significant results become insignificant, indicating the results may be affected by omitted variable bias.

The rest of the paper is organized as follows. Section 2 is a review of previous literature on identity, gender stereotypes surrounding math, and the relationship between gender identity and math performance. Section 3 presents the theoretical framework and hypotheses. Section 4 is a description of the data and methodology. Section 5 contains the results and analysis. Finally, Section 6 concludes.

2. Literature Review

This paper is related to several strands of literature. The first strand is a small but growing literature which explores how identity can affect economic outcomes (e.g., Akerlof and Kranton, 2000; Austen-Smith and Fryer, 2005; Fryer and Torelli, 2010; Bursztyn et al., 2017 and Brenøe et al., 2022). Among the firsts to provide a framework for analyzing how identity may impact behavior and economic outcomes were Akerlof and Kranton (2000). To do so they propose that identity should be included in the utility function. In their model, an individual’s identity is based on belonging to a set of different social categories, e.g., “man” or “woman”, and each category is associated with prescriptions of appropriate behaviors and attributes. Because there are different ideals for men and women certain behaviors and attributes are prescribed for one gender, but not the other. Akerlof and Kranton argue these prescriptions can be internalized, hence violating them may imply both social and internal costs (e.g., to self-image or identity). As an example, they claim a woman working in a male dominated field may suffer a utility loss for violating the prescribed behavior for women. This argument implies not only that identity can affect behavior, but also that behavior and attributes may affect identity.

The argument that behavior can affect identity is strengthened by the findings of Fryer and Torelli (2010). Their results show that, above a certain threshold, minority students in the United States perceive a negative correlation between social status and high grades. They claim black students who invest in education risk being perceived as "acting white" by their peers, and therefore suffer an opportunity cost of decreased social status when they perform well in school. In effort to maintain group loyalty, these students may act in accordance with the prescribed behavior for their social group by not pursuing high grades. Closer to the gender focus in this paper, Bursztyn et al. (2017) find a similar result among single female MBA students regarding high performance and career ambition. Their results show that single female MBA students participate less in class compared to their peers, and when they expect their answers to be observed they state less career ambition compared to their married female classmates. Bursztyn et al. describe this behavior as "acting wife", and they speculate that single female students state less ambition to be perceived as more attractive in the marriage market. Thus, single women may "act wife" by displaying behaviors in line with the prescriptions of behavior for women to fulfill feminine stereotypes and possibly increase their prospects in the marriage market. Bertrand et al. (2015) observe a sharp drop in the distribution of the share of income earned by the wife once the wife earns more than half of the total household income, indicating the fraction of couples where the wife earns more than the husband is much smaller than the fraction of couples where the husband earns more than the wife. They argue this drop is driven by gender identity norms.[2] Their results show that if a randomly selected woman in

---

[2]Whether this drop is caused by gender identity norms is debated. Hederos and Stenberg (2022) find a similar drop in Swedish data, but they do not find strong evidence that it is caused by gender identity norms. Instead, they argue that the large share of couples who earn exactly the same (e.g., couples who run a business together) cause a spike in the distribution at 0.5, which increases the drop. This spike is not visible in Bertrand et al. (2015) because the representation of the distribution in their study is coarser.

the marriage market is likely to receive a higher salary than a randomly selected man, marriage rates decrease. They also find married couples are more likely to be unhappy in their marriage and to get divorced if the wife earns more than the husband. These findings suggest that acting in line with the prescriptions of your gender may increase success in the marriage market.

This paper is also related to a strand of literature, mainly within psychology, which explores gender stereotypes related to math ability and how they may impact performance. For example, according to Cvencek et al. (2011) the stereotype that math is for boys arise at an early age. Using a sample of 247 American children between the ages six to ten, they find that both boys and girls associate math more strongly with boys than with girls. Furthermore, using a sample of 34 countries, including Sweden, Nosek et al. (2009) find a relationship between nation-level implicit stereotypes and sex differences in eight grade math performance. They also find that the difference is larger in countries where math is considered a male subject. Nosek et al. speculate this is one of the reasons there still exists a gender gap in math intensive industries. Nosek et al. (2002) explore the relationship between math stereotypes and identity. They find that women who identify as female and have implicit stereotypes associating math with maleness found it difficult to associate math with themselves. This was true even for women in math intensive academic tracks. They also found that women with these stereotypes showed negative attitudes towards math, while men with these stereotypes exhibited positive attitudes towards math. While most research examining gender stereotypes surrounding math and how they affect performance are within psychology, the topic has been discussed in the economic context as well. Carlana (2019) document the impact of these stereotypes on the gender gap in math performance and educational choice. She finds that girls with teachers who associate math with maleness perform worse on math tests and are less likely to choose math intensive academic tracks compared to girls who are not subject to such stereotype treatment.

The final strand of literature this paper relates to explores the relationship between gender identity and math ability or attitudes towards math. Schmader (2002) found that women who value gender identity strongly perform worse than men on math tests where they know their results will be used as a measure of relative performance between genders. Schmader highlights that the results cannot confirm if it is fear of contributing to gender stereotypes which make these women perform worse, or if it is the result of being primed with gender stereotypes before the test. In a similar vein, McGeown and Warhurst (2020) examine how primary school children's motivation for different school subjects (reading, writing, science, and mathematics) is affected by gender identity. They find boys report greater confidence in math, but they do not find a significant difference between the correlations of masculine and feminine traits and math motivation. The authors claim the perception that math is a male-oriented subject increases with age and argue this could explain why they do not find significant gender differences in math motivation amongst primary school children.

Previous literature discussed in this section explores how gender identity or stereotypes may affect math performance. This paper adds to the literature by exploring the opposite perspective. While it has been found that identity can affect behavior, there are also arguments that have been discussed in this section which suggest behavior can affect identity (e.g., Akerlof and Kranton, 2000; Fryer and Torelli, 2010; Bursztyn et al., 2017). This, combined with the fact that math ability is associated with male stereotypes, makes it plausible to assume perceived high math performance may causally impact gender identity. Therefore, this paper explores a perspective which, to my knowledge, has not been examined before.

3. Theoretical Framework

3.1 Identity as Part of the Utility Function

As mentioned, Akerlof and Kranton (2000) were among the firsts to provide an analytical framework to analyze how identity may impact behavior and economic outcomes. In the proposed framework they propose, an individual's identity is based on belonging to different social categories (e.g., "man" or "woman"). Social norms prescribe the appropriate behaviors and characteristics for each social category, and deviating from these prescriptions is costly. To illustrate how identity may influence behavior Akerlof and Kranton propose the following utility function:

$$U_j = U_j\left(\boldsymbol{a}_j, \boldsymbol{a}_{-j}, I_j\right) \tag{1}$$

They define the utility of individual j as dependent on a vector of j's actions, $\boldsymbol{a}_j$, others' actions $\boldsymbol{a}_{-j}$, and identity I_j. Since $\boldsymbol{a}_j$ and $\boldsymbol{a}_{-j}$ describe the actions of individual j and others, and this determines j's consumption of goods and services, Akerlof and Kranton argue that this utility function also encompasses the standard utility function. They continue to characterize identity as outlined:

$$I_j = I_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}; \boldsymbol{c}_j, \boldsymbol{\epsilon}_j, \boldsymbol{P}) \tag{2}$$

The identity of individual j depends on j's assigned social categories, $\boldsymbol{c}_j$, j's own given characteristics, $\boldsymbol{\epsilon}_j$, and the prescriptions of behavior and characteristics for the different social categories, $\boldsymbol{P}$. Individual j's identity is also affected by how well j's own characteristics and behaviors correspond to the ideal of j's social category, and whether others' characteristics and behaviors correspond to the ideals of their social categories. In the equations above it can be observed that how individual j's actions affect utility partly depends on how they affect identity. This is an example of why men and women can have different payoffs from behaviors stereotypically associated with a specific gender. While identity can be affected by an individual's own actions, it can also be affected by others' actions. If someone acts in a way that goes against individual j's prescribed behavior for that person's social category, this can threaten the identity of j. Akerlof and Kranton exemplifies this by describing that a man's identity might be threatened by another man wearing a dress, or a woman being a Marine.

Because violating the prescriptions of behavior is associated with an identity cost, and therefore a utility loss, people may avoid displaying behaviors and characteristics which are not prescribed for their social category. As math is associated with a male stereotype this implies women may want to avoid being perceived as good at math. Furthermore, women who are good at math may suffer a utility loss for violating the prescriptions of behavior for women.

3.2 Hypotheses

Applying the theoretical framework presented by Akerlof and Kranton (2000) to the topic of this thesis, it implies that women may want to avoid exhibiting stereotypical male qualities like math skills. This argument is strengthened by the findings of Fryer and Torelli (2010) and Bursztyn et al. (2017), which provide evidence that violating prescriptions of behavior can be costly. The framework by Akerlof and Kranton (2000) also implies that people who have internalized the male stereotype surrounding math may identify as more masculine if they perceive themselves as good at math. Thus, women who perceive themselves as good at math may suffer an identity cost for violating the prescriptions of behavior, while men who perceive themselves as good at math may experience an identity boost for conforming to the prescriptions of behavior. I therefore hypothesize that, on average:

1. Individuals who perceive themselves as good at math identify as more masculine.
2. Individuals who perceive themselves as good at math believe others perceive them as more masculine.

Because the hypotheses are built on the assumption that math is prescribed to be a male subject, the effect is expected to be found amongst those who have internalized these prescriptions.

4. Data and Methodology

To test the hypotheses, I conduct an experiment. The first part of this section describes the data and the experiment. The second part defines the empirical approach. Finally, in the third part I discuss potential ethical concerns and limitations to the study.

4.1 Data and Experiment Design

The data was collected using an experiment in the form of a two-part online questionnaire. The questionnaire was sent out via email to 8364 economics, business, and finance students at 17 Swedish universities and was completed by 280 people in total. Optimally, I would have conducted the study on students from many different programs, however this was not possible within the given time frame due to administrative reasons. Because the hypothesized effect has most implications amongst those who could have selected math intensive academic tracks but chose not to, I limited the study to a few programs which contain math but are not extremely math intensive. The questionnaire can be found in full in Appendix A. The first part of the questionnaire consisted of a math test comprising 10 multiple choice questions. The difficulty level was set to equivalate the Swedish Scholastic Aptitude test (SweSAT) or the Swedish mathematics 3b/3c courses. This difficulty level was chosen because it represents the prerequisites for the economics, business, and finance bachelor programs at Swedish universities. After the participants had completed the test the treatment variation was implemented. Participants were randomly assigned to receive performance feedback in relation to either a high performing or low performing reference group. Participants who were compared to the low performing reference group were hence, on average, exogenously assigned to receive a better signal of their relative performance than those compared to the high performing reference group. The formation of the reference groups is outlined in the next section.

After the participants were informed about their relative performance, they were asked two questions about their gender identity. First, the participants were asked to place themselves on an 11-point scale ranging from very masculine to very feminine. Then they were asked to estimate how others would place them on the same scale. The scale has been tested in previous work and has been found to correlate well with more extensive measures of gender identity (Brenøe, et al., 2022). Since it is also more time efficient and straight forward compared to more extensive measures, it is appropriate for this type of experiment. Because asking about gender identity can be sensitive the participants could select to not answer these questions. Only five participants selected this option and were therefore excluded from the analysis.

After the main questions, a set of exploratory questions were asked. First the participants were asked how many days per month they would be willing to travel for work and how many hours per week they would be willing to work after they graduate. These questions were inspired by Bursztyn et al. (2017) and were included to examine if treatment may impact attitudes towards future career and work life. The reasoning is that if treatment has an impact on gender identity this may subsequently affect future career outcomes. Since I cannot measure actual career outcomes, I chose to measure attitudes in certain scenarios. The participants were also asked to estimate their confidence and ambition on 11-point scales ranging from not confident (ambitious) at all to very confident (ambitious). Because the hypotheses are based on the assumption that math is prescribed as a male subject, the participants were also asked to place mathematics on the same masculine to feminine scale as used for the gender identity questions. This question was included to reveal the degree to which a participant associated math with masculinity. The distribution is presented in Figure B1 in Appendix B. After this the participants were asked to rate their own math ability on an 11-point scale ranging from very bad at math to very good at math. This question was asked as a manipulation check to examine whether the treatment had the desired effect on the participants perception of their relative math ability. As a further robustness check, the participants were asked if they remembered how many percent of students from the comparison group they outperformed. The purpose of this question was to control that the participants had read the relative performance feedback. They were also asked what they believed the purpose of the study was. This question was included because if a participant correctly understood the purpose of the study, it might have affected their answers. At the end of the questionnaire the participants were asked a few socio-demographic questions. These were legal gender, age, which program they are enrolled in and at which level they are studying.

4.1.1 The Reference Groups

The reference groups comprised 100 economics, industrial economics, marketing, and management students from three Swedish universities. These programs were chosen because they are related to the economics and business fields, which are the target groups for the study, but the math intensiveness differ between the programs. Economics, and especially industrial economics, are more math intensive while marketing and management are less math intensive. Hence, the students were divided into two groups, with marketing and management students in one and economics and industrial economics students in the other. In total, there were 19 people in the marketing and management group and 81 in the economics and industrial economics group.[3] Out of the 81 people in the economics group, 71 were industrial economics students. The economics/industrial economics group scored an average of 8.49 correct answers, and the marketing and management group scored an average of 4.37 correct answers. Hence, the economics group was used as the high performing reference group, and the marketing and management group was used as the low performing reference group. Figures B2 and B3 in Appendix B present the performance distribution in the two groups. It would have been nice if the distribution of scores in each reference group had been a bit smoother, but the fact that the average scores significantly differ is what is most important.

4.2 Empirical Approach

In this section the main regression models and variables are specified. An overview of all variables used can be found in Table B1 in Appendix B. To test the hypotheses, I begin by running OLS regressions using first and second order gender identity as dependent variables, and the treatment indicator as the independent variable. Gender identity, *GI*, is how the participants perceive their own gender identity and is measured on a scale ranging from "Very masculine" = 1 to "Very feminine" = 11. The participants' beliefs about how others perceive their gender identity, GI_{Others}, is measured on the same scale. Treatment, *HighTreatment*, is a binary variable, where 1 indicates the participant received performance feedback signaling high relative performance, i.e., their score was compared to the low performing reference group, and 0 indicates the participant received performance feedback signaling low relative math performance, i.e., their score was compared to the high performing reference group. Since I am conducting an experiment where treatment is randomized, my main specification does not include additional control variables. However, since the randomization of treatment was not

[3] The reason there is an uneven distribution between the two reference groups is that there was a much higher response rate from the industrial economics program compared to any other program.

stratified by gender, and more women (from less math intense programs) were randomly assigned to one treatment, I also run specifications controlling for this at the final stage of analysis.

$$GI_i = \beta_0 + \beta_1 HighTreatment_i + \varepsilon_i \quad (3)$$

$$GI_{Others_i} = \beta_0 + \beta_1 HighTreatment_i + \varepsilon_i \quad (4)$$

I first run models 3 and 4 specified above using the full sample, and then on men and women separately. Then, because the hypotheses rely on participants associating math with masculinity, I will also run the regressions excluding those who have stated that they consider math a feminine subject. Finally, to analyze if the effect differs between men and women, I will run the regressions with an interaction between treatment and gender. In these regressions I include a binary variable for legal gender, *Male*, where male = 1 and female = 0. This considerably lowers statistical power and will therefore be viewed as exploratory analysis.

$$GI_i = \beta_0 + \beta_1 HighTreatment_i + \beta_2 Male_i + \beta_3 HighTreatment_i \times Male_i + \varepsilon_i \quad (5)$$

$$GI_{Others_i} = \beta_0 + \beta_1 HighTreatment_i + \beta_2 Male_i + \beta_3 HighTreatment_i \times Male_i + \varepsilon_i \quad (6)$$

In equation 3, β_1 is interpreted as the effect the high treatment (receiving feedback indicating high relative math performance) has on gender identity, on average. In equation 4, β_1 is interpreted as the effect treatment has on the participant's beliefs of how others perceive their gender identity. In equation 5, β_1 is interpreted as the effect treatment has on gender identity for women, and β_1+ β_3 is interpreted as the effect treatment has on gender identity for men. If β_3 is significantly different from zero, I will be able to conclude that the effect treatment has on gender identity is significantly different for men and women. β_2 is the difference between men and women in the low treatment condition. This coefficient is of less importance to my study since I expect on average men will identify as more masculine than women regardless of which treatment they receive. Finally, the coefficients in equation 6 are interpreted in the same way as in equation 5, but with GI_{Others} as the dependent variable instead.

4.3 Ethical Concerns and Limitations

A potential ethical concern is that I ask about gender identity, a subject which some participants might find sensitive. As mentioned, to avoid making people answer questions they may find too personal or intimate, the participants were provided with an option to not answer the gender identity questions. Because the questions about gender identity are the dependent variables, the five participants choosing not to answer this question are excluded from the analysis. A number of limitations are also worth mentioning. One concern with this type of study is that participants might understand what the purpose of the study is, and that such insights might have affected their answers. For this reason, I asked the participants what they believed the purpose of the study was. Only seven participants gave answers indicating they may have understood the purpose of the study, and dropping these individuals does not significantly impact the results. Another limitation is that it is possible people who like math were more willing to complete the questionnaire, which implies the sample may not accurately represent the population. A particular problem is if women who selected into the study like math, and if these women experience a different relationship between gender identity and math performance than other populations, this may have dampened the results. Finally, and maybe most importantly, I am looking for an effect of a small manipulation on a self-reported and possibly noisy measure. Participants' perception of their mathematical aptitude is based on a long history of learning math, and my treatment variation is likely to have at most a minor effect on this (which is also a reason for why I include a manipulation check). Moreover, there are arguably several determining factors for gender identity, which means it is possible that a small manipulation of perceived math performance will not yield a large effect. Adding to this the fact that the sample is rather small, my study likely has low power. This implies I may not be able to find an effect even if there is one, and interpretation of the results must be made with caution. Therefore, this should only be viewed as a first study which I would like to develop it in the future. A central development for a future study will be to find a treatment with a large enough impact.

5. Results and Analysis

5.1 Descriptive Statistics

In total 280 participants completed the questionnaire. After excluding the observations of those who did not answer the gender identity questions, or said they were not university students, 270 observations remained.

Table 1.
Descriptive Statistics.

Gender	High Treatment	Low Treatment	Total
Men	76	74	150
Women	53	67	120
Age			
<20	11	8	19
20-25	94	118	212
26-30	18	10	28
31-35	3	4	7
36-40	3	0	3
41-45	0	1	1
Program			
Business	46	56	102
Economics	25	24	49
Business & Economics (in Swedish Civilekonom)	39	43	82
Finance	7	5	12
Other Program (within Business/Economics/Finance)	12	9	21
Other Program (not within Business/Economics/Finance)	0	4	4
Current Level of Education			
Bachelor – Year 1	46	44	90
Bachelor – Year 2	35	47	82
Bachelor – Year 3	44	45	89
Master – Year 1	3	5	8
Master – Year 2	1	0	1
Score			
Mean	5.527	5.354	5.437
Median	6	5	5
Total	**129**	**141**	**270**

There are no major differences between the treatment groups, but one concern is that the proportion of men and women differs slightly. Since the hypothesis is that participants receiving the high treatment will perceive themselves as more masculine, the fact that there are relatively more men in that group may cause spurious results. There is also a difference in proportions of business students between the treatment groups. To control if the differences are significant, I run proportions tests. The difference between proportion of men and women in the high treatment group (*M* = 0.589) and the low treatment group (*M* = 0.525) is insignificant, *z* = -1.063, *p* = 0.288. Similarly, the difference between proportion of business students in the high treatment group (*M* = 0.357) and the low treatment group (*M* = 0.397) is also insignificant, *z* = 0.687, *p* = 0.492. Even though the proportions tests show the differences in proportions are insignificant, because gender is highly correlated with gender identity, I will control for this in the last step of the analysis.

Another potential problem is that the average score is slightly higher in the high treatment group. I therefore run a two-sample t-test. Although the high treatment group has a higher average score (*M* = 5.527, *SD* = 2.550) than the low treatment group (*M* = 5.355, *SD* = 2.367) this difference is statistically insignificant, *t*(268) = -0.577, *p* = 0.565. I also control if the average score differs between treatment groups for men and women respectively. The average score is higher for men in the high treatment group (*M* = 6.224, *SD* = 2.474) than in the low treatment group (*M* = 5.757, *SD* = 2.426), but it is not statistically significant, *t*(148) = -1.167, *p* = 0.245. For the women the average score is higher in the low treatment group (*M* = 4.910, *SD* = 2.234) than in the high treatment group (*M* = 4.528, *SD* = 2.334) but this difference is not statistically significant either, *t*(118) = 0.912, *p* = 0.364. It therefore seems to be no major differences between the two groups apart from which performance feedback they received.

Since men have higher average scores than women in both treatment groups, I run a t-test to control if the difference in score between genders is significant. I find that the men have a higher average score (*M* = 5.993, *SD* = 2.454) than the women (*M* = 4.742, *SD* = 2.277), and that the difference is statistically significant, *t*(268) = -4.300, *p* = 0.000. Since the average scores of men and women separately does not differ between treatment groups, the fact that the women on average scored lower than the men should not be a problem. However, because implications may be largest for women who are good at math, it is unfortunate the sample does not contain more high performing women. The score distributions are presented in Figures B4 and B5 in Appendix B.

5.2 Main Results and Analysis

I begin by plotting the distributions of the participants' own perception of gender identity in the different treatment conditions for the full sample, and for men and women separately. The graphics for participants beliefs of how others perceive their gender identity are very similar to those for own perception of gender identity, hence I do not present them in this section. They can be found in Figures B6-B8 in Appendix B.

There does not seem to be a big difference between the treatment groups in the full sample. If Hypothesis 1 is true, the distribution of the high treatment group should be shifted towards masculine compared to the low treatment group. A two-sample t-test confirm that even though the high treatment group (*M* = 5.620, *SD* = 3.088) has a lower mean value, indicating the mean is shifted towards masculine, than the low treatment group (*M* = 5.929, *SD* = 2.961) there is no significant difference in perception of own gender identity between the two treatment groups, $t(268) = 0.839$, $p = 0.402$.

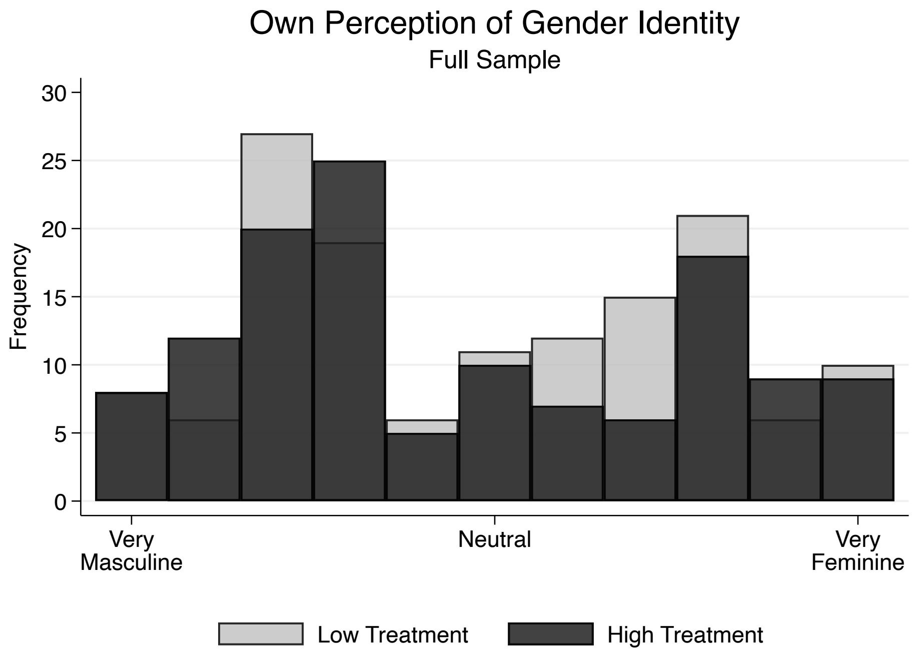

Figure 1. *GI Distribution - Full Sample.* Notes: Low treatment indicates receiving feedback of low relative math performance. High treatment indicates receiving feedback of high relative math performance.

When plotting the observations for women separately there still does not seem to be an evident difference between the groups. If anything, it seems women in the low treatment group tends to place themselves towards masculine compared to the women in the high treatment group. When doing a t-test I find no significant difference, $t(118) = -0.510$, $p = 0.611$, between the low treatment group (M = 8.478, SD = 1.735) and the high treatment group (M = 8.642, SD = 1.766).

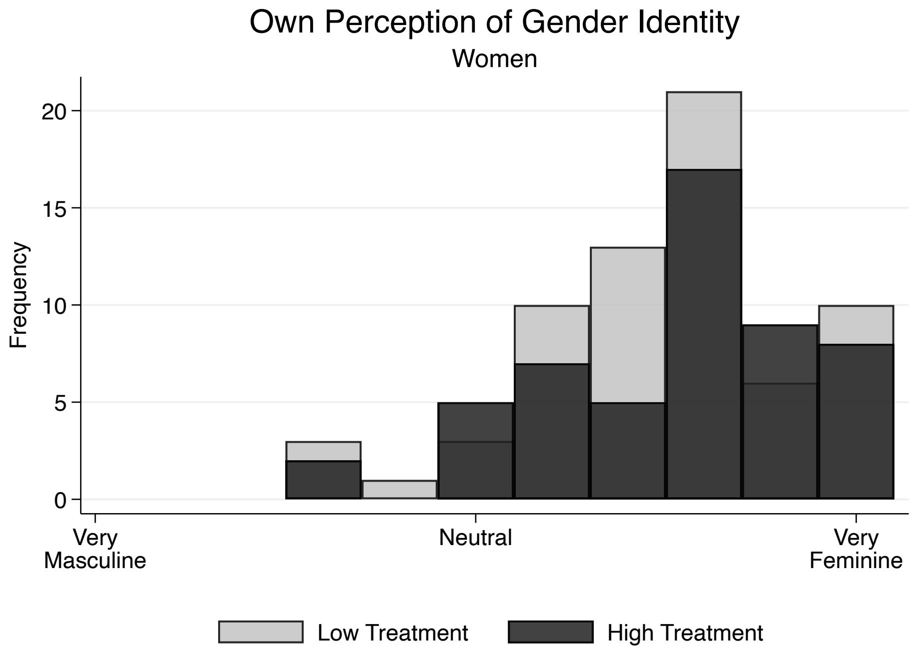

Figure 2. *GI Distribution - Women.* Notes: Low treatment indicates receiving feedback of low relative math performance. High treatment indicates receiving feedback of high relative math performance.

There is also no obvious difference between treatment groups for men. The t-test confirm that the low treatment group (*M* = 3.622, *SD* = 1.653) and the high treatment group (*M* = 3.513, *SD* = 1.778) does not differ significantly, $t(148) = 0.387$, $p = 0.700$.

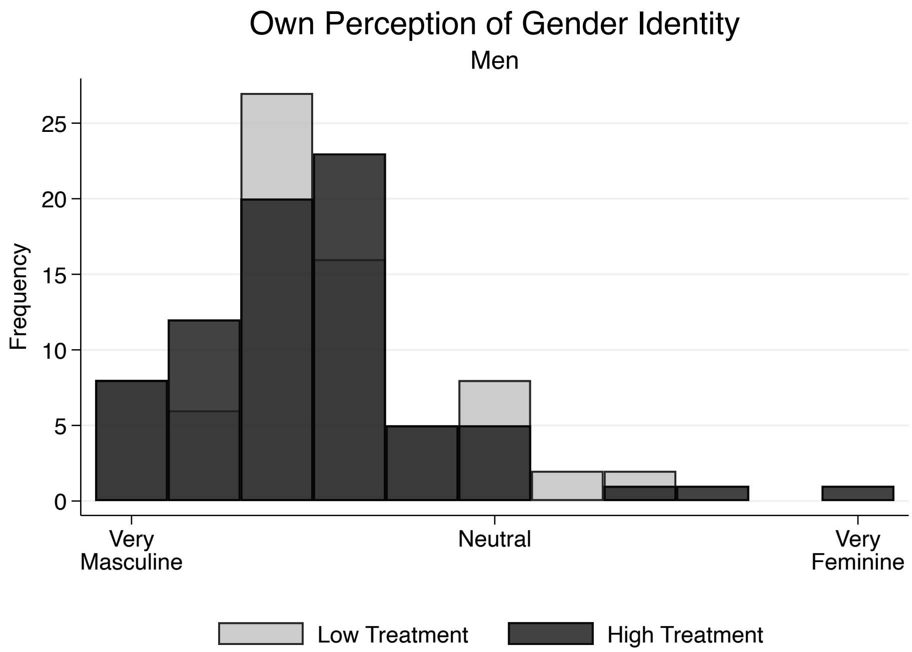

Figure 3. *GI Distribution - Men.* Notes: Low treatment indicates receiving feedback of low relative math performance. High treatment indicates receiving feedback of high relative math performance.

Moving on to the regression analysis, first OLS-regressions are performed with an indicator variable for high treatment as the independent variable and gender identity and the participants' beliefs about others' perception of their gender identity as the dependent variables. These regressions are run on the full sample and for men and women separately. Because I included a control question to reveal potential stereotypes the participants had surrounding math, and the effect is likely to be largest amongst those who have either explicit or implicit stereotypes associating math with masculinity, I also ran the regressions once dropping the observations of those who indicated they associated math with femininity. Amongst the 270 participants, 10% associated math with femininity, 43.3% associated math with masculinity and 46.7% indicated they view math as gender neutral. To maintain a decent sample size, I keep the observations of those who consider math gender neutral for now. However, in the next section I run the regressions only including the observations of those who stated they view math as masculine as a robustness check. The distribution of the participants stereotypes is presented in Figure B1 in Appendix B.

Table 2.

Treatment on Gender Identity (GI).

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.309 (0.369)	-0.108 (0.280)	0.164 (0.322)	-0.764** (0.376)
Constant	5.929*** (0.249)	3.622*** (0.192)	8.478*** (0.212)	5.923*** (0.257)
N	270	150	120	243
R^2	0.003	0.001	0.002	0.017

Notes: Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table 2 shows the result of regressing high treatment on gender identity for the full sample in column 1, for men and women respectively in columns 2 and 3, and for the full sample excluding those who considered math feminine in column 4. For the full sample and for men and women respectively the coefficient for treatment is not significant, and Hypothesis 1 cannot be confirmed. However, when excluding the observations of those who indicated they had feminine stereotypes surrounding math, the coefficient for treatment is negative and statistically significant at the 5% level. For this group, treatment is associated with a shift towards masculine on the gender identity scale by 0.764 points. Given that the gender identity scale only has 11 points, and that the treatment may only have had a minor effect on the participant's perception of their mathematical aptitude, a shift of 0.764 in the hypothesized direction is rather substantial. While the test in column 4 is the one I believe is most appropriate since I expect the effect to be found amongst those who associate math with masculinity, it is only one test out of four which yields a significant result. Thus, the results need to be interpreted cautiously, and may only be seen as an indication along the lines of my hypothesis.

Table 3.

Treatment on Beliefs About Others' Perception of Gender Identity (GI_{Others}).

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.329 (0.358)	-0.060 (0.274)	0.008 (0.359)	-0.810** (0.364)
Constant	5.965*** (0.247)	3.770*** (0.185)	8.388*** (0.246)	5.969*** (0.259)
N	270	150	120	243
R^2	0.003	0.000	0.000	0.020

Notes: Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table 3 shows the same regression as Table 2 but with the participants beliefs of others' perception of their gender identity as the dependent variable. Again, the coefficient for high treatment is insignificant in the first three columns. When those who consider math to be a feminine subject are excluded, the coefficient for treatment is negative and statistically significant at a 5% level. Treatment is associated with a 0.810-point shift towards masculine on the scale measuring beliefs of how others perceive their gender identity. As before, within the context of my study, this is a substantial shift in the hypothesized direction. Like the results in Table 2, there is only one out of four tests which yield a significant result, meaning interpretations should be made cautiously.

To control for gender differences, I run the regressions while including an interaction term between high treatment and legal gender. The aim is to detect if the treatment effect differs between men and women. As mentioned above, this considerably lowers power and should be viewed as exploratory analysis.

Table 4.

Treatment on GI - Gender Differences.

	Full Sample (1)	Excl. Feminine Stereotypes (2)
High Treatment	0.164 (0.322)	-0.196 (0.334)
Male	-4.856*** (0.286)	-4.920*** (0.281)
High Treatment × Male	-0.272 (0.427)	-0.189 (0.423)
Constant	8.478*** (0.212)	8.610*** (0.204)
N	270	243
R^2	0.675	0.716

Notes: Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

In table 4 the coefficients for high treatment are insignificant for the full sample and the full sample excluding those who consider math to be a feminine subject. The coefficient for the interaction term is also insignificant in both columns. It therefore seems the effect of treatment does not differ between genders, but as mentioned this test has lower statistical power.

Table 5.

Treatment on GI_{Others} - Gender Differences.

	Full Sample (1)	Excl. Feminine Stereotypes (2)
High Treatment	0.008 (0.359)	-0.552 (0.387)
Male	-4.618*** (0.308)	-4.773*** (0.313)
High Treatment × Male	-0.068 (0.452)	0.277 (0.465)
Constant	8.388*** (0.246)	8.576*** (0.249)
N	270	243
R^2	0.622	0.653

Notes: Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

The coefficients for high treatment and the interaction term are still statistically insignificant in both columns when the dependent variable is beliefs of how others perceive their gender identity, indicating the treatment effect does not differ between genders.

Even though the proportions test did not show a significant difference in the proportion of men and women between the two treatment groups, the fact that there are relatively more men in the high treatment group will likely cause the average perception of gender identity in this group to be more shifted towards masculine. Since the hypothesis is that treatment will shift the average in the same direction, not controlling for gender may yield spurious results. Since there were also some differences in the proportion of business students between the groups, I run the analysis once including gender and academic program as control variables[4].

Table 6.

Treatment on GI & GI_{Others}: Controlling for Gender and Program.

	GI		GI_{Others}	
	Full Sample (1)	Excl. Feminine Stereotypes (2)	Full Sample (3)	Excl. Feminine Stereotypes (4)
High Treatment	0.026 (0.218)	-0.317 (0.214)	-0.003 (0.228)	-0.376 (0.228)
Male	-5.010*** (0.217)	-5.033*** (0.214)	-4.643*** (0.235)	-4.666*** (0.237)
Economics	-0.040 (0.260)	0.078 (0.256)	-0.412 (0.325)	-0.182 (0.331)
Business & Economics	0.080 (0.282)	-0.111 (0.266)	-0.159 (0.275)	-0.245 (0.269)
Finance	0.219 (0.514)	0.471 (0.529)	-0.391 (0.533)	0.165 (0.403)
Other Related Program	-0.058 (0.380)	0.022 (0.383)	-0.231 (0.415)	-0.152 (0.469)
Other Unrelated Program	0.487** (0.240)	0.366 (0.246)	0.355 (0.3222)	0.297 (0.330)
Constant	8.521*** (0.234)	8.667*** (0.235)	8.539*** (0.237)	8.620*** (0.245)
N	270	243	270	243
R^2	0.675	0.718	0.626	0.654

Notes: Dependent variable is Gender Identity (GI) in columns 1 & 2, and Others' Perception of Gender Identity (GI_{Others}) in columns 3 & 4. Both are measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Business is used as reference for programs. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

[4] I also run this regression controlling for actual performance (i.e., score) and it does not affect the main results.

When controlling for gender and program there is no significant effect of treatment on first or second order gender identity, even when excluding those with feminine stereotypes. These results indicate the significant results in Tables 2 and 3 may be spurious, and hence I cannot conclude there is a treatment effect. However, the coefficients for treatment still have the hypothesized signs in column 2 and 4, and even though the magnitude of the coefficients have decreased the effect is still rather substantial if it had been significant. Since the sample is small, and the treatment likely only has a minor effect on perceived math ability, it is possible there is an effect even though I have not been able to confirm it.

5.3 Manipulation and Robustness Checks

To confirm that receiving the high treatment made participants perceive their math performance as relatively high, as a manipulation check the participants were asked to estimate their math ability. For the full sample, receiving the high treatment is associated with a positive point estimate of approximately 0.5, indicating treatment tends to be associated with a 0.5-point increase in estimated ability. However, the difference fails to reach the conventional significance level ($p = 0.058$). When running the regression for men and women separately, high treatment is associated with a positive increase in estimated ability for both genders, but the results are only significant at the 10% level for men, and insignificant for women. The results from regressing high treatment on estimated math ability can be found in Table B2 in Appendix B. The manipulation check indicates that treatment was not strong enough to significantly alter the participants' perception of their math ability, which may explain why there were no conclusive significant results of treatment on gender identity.

As robustness checks, the regressions in Tables 2-5 were run excluding those who could not remember their performance feedback and those who understood the purpose of the study. These results show that the treatment is significant at the 5% level when excluding those with feminine stereotypes, but not for the full sample or men and women separately. There is still no evidence the treatment effect differs between genders. These tests are found in Tables B3-B10 in Appendix B. To examine the robustness of the results received when controlling for stereotypes, I run the tests once only including only those who consider math masculine. These results show a larger negative effect of treatment on gender identity than when only excluding those with feminine stereotypes, but the difference is only significant at a 10% level. The results are found in Tables B11 and B12 in Appendix B. Running the regressions on only those who consider math to be masculine drastically decreases the number of observations, and the loss of statistical power may explain why the results become less significant than the results received when also including those who view math as gender neutral. Furthermore, it is possible that some participants who have internalized the male stereotype surrounding math indicated they view math as gender neutral, either because they do not realize they subscribe to the stereotype or because they do not want to subscribe to the stereotype. Considering the distribution of stereotypes indicate there is a male stereotype surrounding math, yet there is a large spike amongst those who indicate they consider math to be gender neutral, this might be the case.

5.4 Exploratory Results and Analysis

As a final part of the analysis, I explore how treatment affects attitudes towards future work life and personality traits. I also examine the correlation between these variables and gender identity.

5.4.1 Treatment Effects on Work Attitudes and Traits

First, the participants were asked about willingness to travel for work and working long hours. These questions were inspired by Bursztyn et al. (2017).

Table 7.

Treatment on Work Attitudes.

	Willingness to Travel for Work			Hours Willing to Work		
	Full Sample	Men	Women	Full Sample	Men	Women
	(1)	(2)	(3)	(3)	(4)	(5)
High Treatment	1.097	2.592**	-0.952	1.350	1.113	0.727
	(0.947)	(1.271)	(1.411)	(1.431)	(1.947)	(1.989)
Constant	9.298***	9.014***	9.612***	45.766***	48.689***	42.537***
	(0.599)	(0.843)	(0.856)	(1.068)	(1.557)	(1.352)
N	270	150	120	270	150	120
R^2	0.005	0.027	0.004	0.003	0.002	0.001

Notes: Dependent variable is Willingness to Travel (days per month) in columns 1-3 and Hours Willing to Work (per week) in columns 3-5. High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

In Table 7 the work attitudes are presented. The treatment does not significantly affect the number of days participants are willing to travel for work in the full sample (column 1) or for women (column 3). However, for the men in the sample (column 2) a signal of high relative math performance is associated with a significant increase in willingness to travel for work by 2.6 days per month. Whether the increase is caused by treatment directly, or indirectly because of how treatment affected gender identity, is unknown. It is possible that receiving a signal of relatively high math performance made men feel more confident or ambitious, and therefore became more willing to travel for work. It is also possible that treatment made men identify as more masculine, and that this resulted in them being willing to work longer hours. Treatment had no significant effect on hours willing to work for the full sample or for men and women separately.

Table 8.
Treatment on Confidence and Ambition.

	Confidence			Ambition		
	Full Sample	Men	Women	Full Sample	Men	Women
	(1)	(2)	(3)	(3)	(4)	(5)
High Treatment	0.381	0.171	0.551	0.010	-0.037	0.111
	(0.267)	(0.347)	(0.403)	(0.255)	(0.341)	(0.384)
Constant	7.092***	7.486***	6.657***	8.021***	7.919***	8.134***
	(0.190)	(0.233)	(0.230)	(0.174)	(0.254)	(0.235)
N	270	150	120	270	150	120
R^2	0.008	0.002	0.004	0.000	0.000	0.001

Notes: Dependent variable is Confidence in 1, 2 & 3 and Ambition 3, 4 & 5. Confidence is measured on a scale from 1 = "Not Confident at All" to 11 = "Very Confident". Ambition is measured on a scale from 1 = "Not Ambitious at All" to 11 = "Very Ambitious". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

In Table 8 I find no significant effect of treatment on the participants confidence or ambition for the full sample or for men and women respectively. Because treatment did not significantly affect confidence or ambition, the significant result in the second column of Table 7 is likely not caused by treatment making men feel more confident or ambitious.

5.4.2 Correlations

As a final part of analysis, I want to explore the correlation between gender identity and math ability and the exploratory variables presented in the previous section. Firstly, the manipulation check indicated the treatment may not have substantially altered the participants perception of their math ability. Even though the participants were not informed about how many correct answers they had, it is plausible participants suspected how well they performed based on how many questions they knew how to solve or how often they just guessed. Furthermore, someone who knows they have always received low grades in math may not believe a positive performance feedback, and vice versa. I will explore the relationship between math ability and gender identity by using actual score and estimated math ability as proxies for math ability. Since I cannot control for potential omitted variables that may correlate with ability, only the correlational relationship will be explored within the scope of this paper.

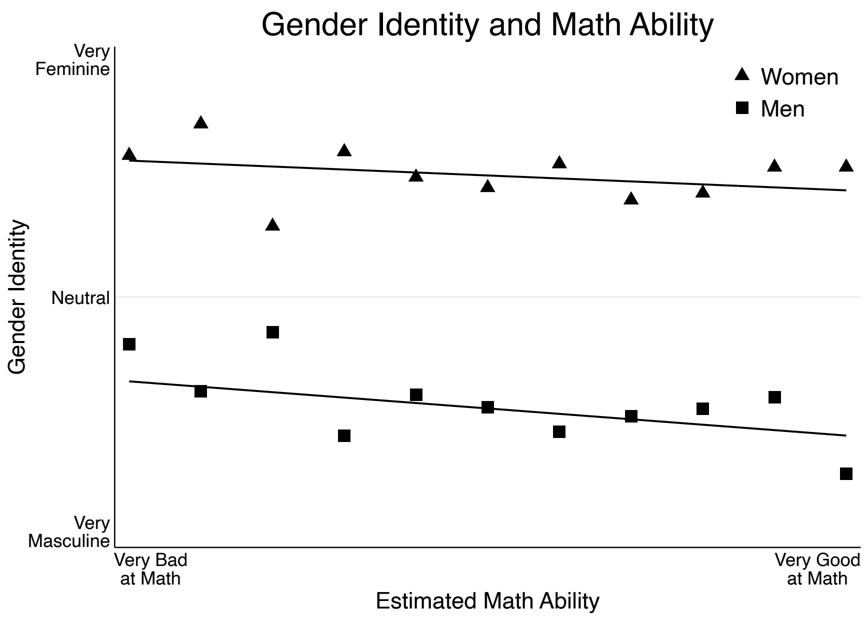

Figure 4. *Correlation Gender Identity and Estimated Math Ability.*

Based on Figure 4 it seems higher estimated math ability tends to be associated with a shift towards masculinity on the gender identity scale for both men and women. The relationship seems to be stronger for men than it is for women, but high estimated math ability is associated with perceiving oneself as more masculine, or less feminine, for both genders. This means the relationship goes in the hypothesized direction. While I cannot confirm causality based on this relationship, I believe it motivates the importance of further research on this topic. As an additional proxy for math ability, I will also look at the relationship between gender identity and score.

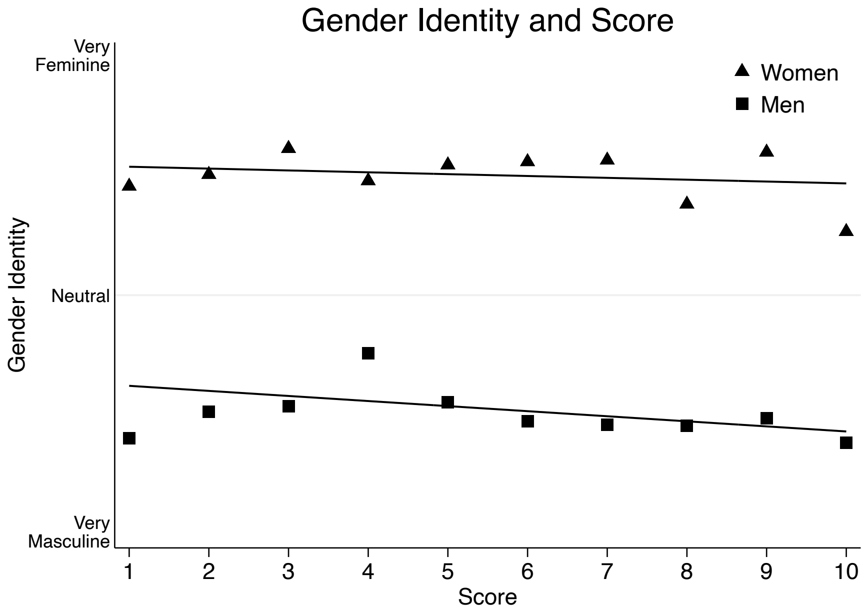

Figure 5. *Correlation Gender Identity and Score.*

While a high score is also associated with a shift towards masculine for both genders, it does not seem to be as strong as the relationship between estimated math ability and gender identity. This implies the participants perception of their own math ability is more highly correlated with gender identity than actual math ability is. Since the participants were not informed about their actual score, it seems reasonable that perceived math ability is more correlated with gender identity than score is. As in Figure 4 above, the relationship is stronger for men.

As a final part of the exploratory analysis, I want to look at the correlations between gender identity and the work attitudes and traits discussed in the previous section. I begin by plotting the relationship for the work attitudes.

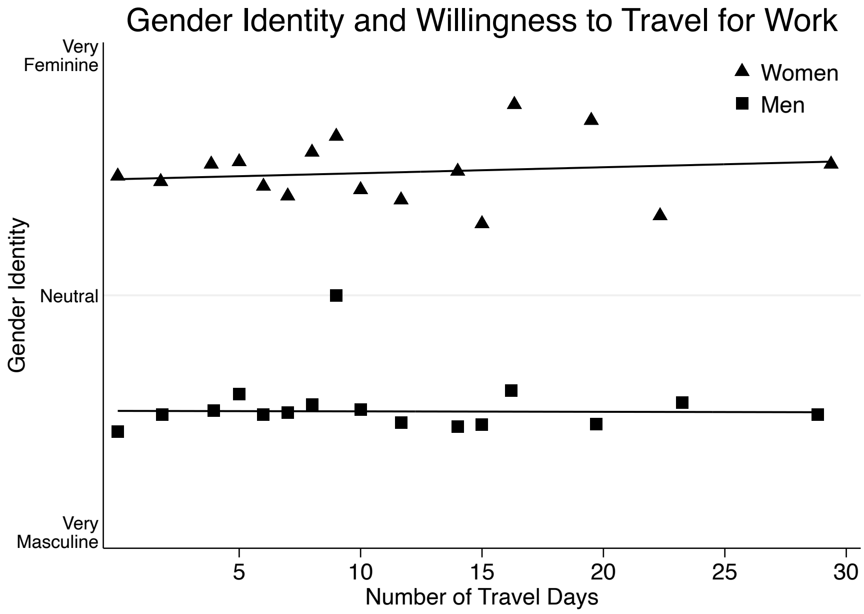

Figure 6. *Correlation Gender Identity and Days Willing to Travel for Work.*

There is no clear relationship between willingness to travel for work and gender identity for men. This indicates that the significant treatment effect on willingness to travel for work for men may not be driven by how treatment affected gender identity. For women there seems to be a small positive relationship, meaning willingness to travel for work is associated with a small shift towards feminine.

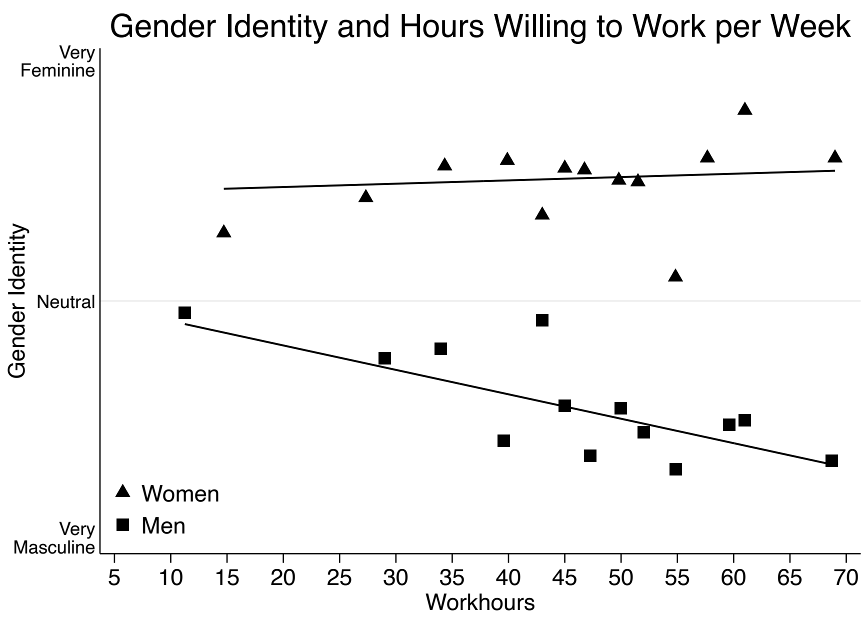

Figure 7. *Correlation Gender Identity and Hours Willing to Work.*

.

There is a strong relationship between gender identity and hours willing to work for men, indicating that being willing to work more hours is associated with a shift towards masculine. I do not want to make assumptions about what is driving this relationship, but it would be interesting to causally explore the relationship and what implications it may have for the labor market. There seems to be a small positive relationship between these variables for women, but it is not evident.

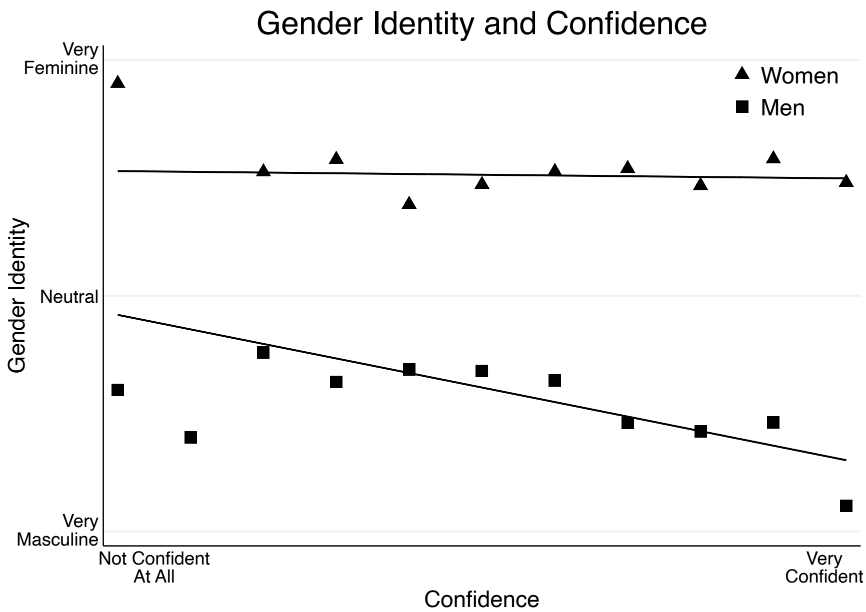

Figure 8. *Correlation Gender Identity and Confidence.*

For women, there is not a strong relationship between gender identity and confidence. However, for men confidence highly associated with a shift towards masculine. This may imply that men who are comfortable with their assigned gender category tend to have higher self-esteem. I find it noteworthy that confidence and gender identity is more strongly linked for men than for women, and it would be interesting to explore implications of this.

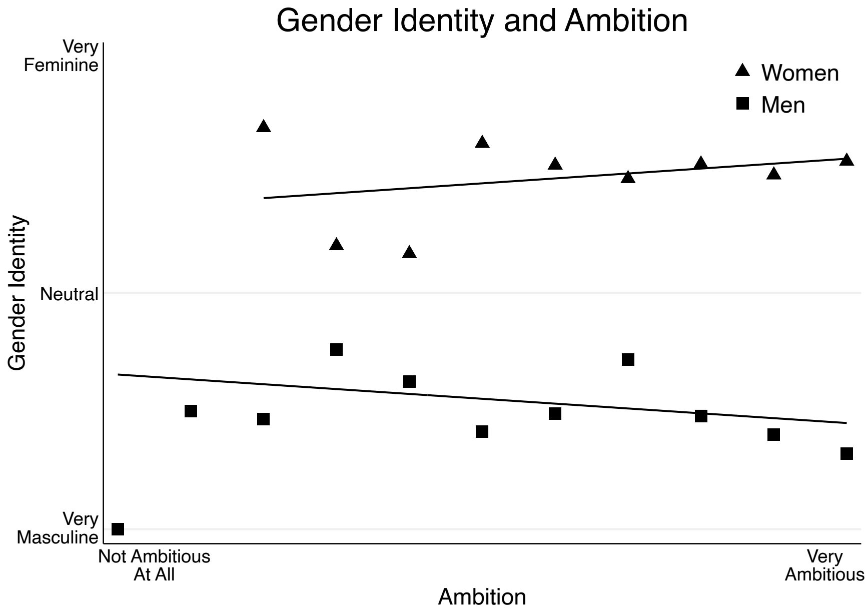

Figure 9. *Correlation Gender Identity and Ambition.*

Finally, gender identity and ambition seem to be correlated for both men and women. For women high ambition is associated with a shift towards feminine, and for men high ambition is associated with a shift towards masculine. Hence, individuals who are comfortable in the gender category assigned to them tend to perceive themselves as more ambitious.

The findings in this section suggest there is a correlation between perceived math ability and gender identity, and that for men gender identity is correlated with work attitudes and traits such as willingness to work long hours and confidence. While these are only correlational findings, I believe they motivate future research on these topics.

6. Conclusion

In this paper I examine the effect of an exogenous signal indicating relatively high or relatively low math performance on gender identity. Because math is generally considered to be associated with male stereotypes, I hypothesized that individuals who perceive themselves as relatively good at math identify as more masculine and believe they are perceived as more masculine by others. I find no significant effect of a treatment signaling high relative math performance on gender identity for the full sample, or for men and women separately. Since the framework this paper is built on suggest the effect should only be found amongst those who associate math with masculinity, I also run the tests excluding the observations of those who considered math feminine. For this group I find a statistically significant treatment effect, indicating that an exogenous signal of high relative math performance tends to make individuals associating math with male stereotypes identify as more masculine. These individuals also believed they were perceived as more masculine by others. I find no evidence that the effect differs between men and women.

Because randomization to treatment was not stratified by gender or program, there ended up being slight differences in the compositions of the treatment groups. Therefore, I also run the tests including gender and program as control variables. With these controls, amongst those who consider math masculine or gender neutral, the magnitude of the treatment coefficient is still substantial and going in the hypothesized direction, however these results are not statistically significant. This indicates the previous results may be affected by omitted variable bias, and therefore the hypotheses cannot be confirmed. Since the treatment only seems to have had a minor impact on the participants perception of their mathematical aptitude, and the sample is small, the study likely has low power. Therefore, the results need to be confirmed in future research.

This has been a first step in exploring how high perceived math performance may impact gender identity. Even though I cannot confirm there was an effect of treatment, I find there is a correlation between perceived math ability and gender identity indicating those who perceive themselves as good at math tend to identify as more masculine. An important next step would be to examine how the gender stereotype surrounding math may affect real outcomes, e.g., choice of academic field. In future research I would therefore like to explore if the hypothesized effect exists amongst high school students, and if it does, how that impacts their choice of tertiary education. It would also be interesting to conduct a similar experiment with a typical female task and see if treatment would result in a gender identity cost for men. The exploratory analysis revealed treatment increased men's willingness to travel for work and that men experience a correlation between hours willing to work and masculinity. These topics would also be interesting to causally explore in future research.

Finally, if women experience a gender identity cost (i.e., identify as less feminine) when working in math intensive fields like STEM, this cost may push women at the margin to select different fields of work. If high skilled women who are suited to work in STEM opt out of these careers as a result of this cost, a consequence may be suboptimal allocation of talent in society. I believe this, combined with the fact that the uneven gender distribution in math intensive fields contributes to the gender wage gap, motivates further research on this topic.

References

Akerlof, G. A. & Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3), pp. 715-753.

Austen-Smith, D. & Fryer, R. G. (2005). An Economic Analysis of "Acting White". *The Quarterly Journal of Economics*, 120(2), pp. 551-583.

Bertrand, M., Kamenica, E. & Pan, J. (2015). Gender Identity and Relative Income Within Households. *The Quarterly Journal of Economics*, 130(2), pp. 571-614.

Blau, F. D. & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature,* 120(2), pp. 551-583.

Brenøe, A. A., Heursen, L., Ranehill, E. & Weber, R. A. (2022). Continuous Gender Identity and Economics. *AEA Papers and Proceedings*, 112, pp. 573-577.

Bursztyn, L., Fujiwara, T. & Pallais, A. (2017). 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *The American Economic Review*, 107(11), pp. 3288-3319.

Carlana, M. (2019). IMPLICIT STEREOTYPES: EVIDENCE FROM TEACHERS' GENDER BIAS. *The Quarterly Journal of Economics*, 134, pp. 1163-1224.

Ceci, S. J. (2018). Women in Academic Science: Experimental Findings From Hiring Studies. *Educational Psychologist*, 53(1), pp. 22-41.

Cvencek, D., Meltzoff, A. N. & Greenwald, A. G. (2011). Math-Gender Stereotypes in Elementary School. *Child Development*, 82(3), pp. 766-779.

Fryer, R. G. & Torelli, P. (2010). An Empirical Analysis of "Acting White". *Journal of Public Economics*, 94(5), pp. 380-396.

Hederos, K. & Stenberg, A. (2022). Gender identity and relative income within households: evidence from Sweden. *The Scandinavian Journal of Economics*, 124(3), pp. 744-772.

Joensen, J. S. & Nielsen, H. S. (2016). Mathematics and Gender: Heterogeneity in Causes and Consequences. *The Economic Journal (London)*, 126(593), pp. 1129-1163.

McGeown, S. P. & Warhurst, A. (2020). Sex differences in education: exploring children's gender identity. *Educational psychology (Dorcherster-on-Thames)*, 40(1), pp- 103-119.

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), pp. 44-59.

Nosek, B. A. et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences – PNAS*, 106(26), pp. 10593-10597.

OECD (2023). *Gender wage gap (indicator).* doi: 10.1787/7cee77aa-en Available: https://data.oecd.org/earnwage/gender-wage-gap.htm#indicator-chart [Accessed 2023-04-06]

Schmader, T. (2002). Gender Identification Moderates Stereotype Threat Effects on Women's Math Performance. *Journal of Experimental Social Psychology*, 38(2), pp. 194-201.

UKÄ. (n.d.). *Antal examinerade 2021*. https://www.uka.se/vara-resultat/statistik/hogskolan-i-siffror/statistik?statq=https://statistik-api.uka.se/api/totals/102 [Accessed 2023-04-13]

UKÄ & SCB. (2023). *Universitet och högskolor. Studenter och examinerade på grundnivå och avancerad nivå 2021/22*. UF 20 SM 2301. Available: https://www.scb.se/contentassets/902b2d1f690a42aab19ba701c2c177db/uf0205_2021122_sm_uf20sm2301.pdf [Accessed 2023-04-13]

Appendix A: Questionnaire

Question 1.

There are two quantities, where:

Quantity I: x
Quantity II: 0

If $x \neq 0$,
And $x^2 = -3x$

Then which of the below is true?

☐ Quantity I is bigger than quantity II

☐ Quantity II is bigger than quantity I

☐ Quantity I is equal to quantity II

☐ The given information is insufficient.

Question 2.

$\frac{1}{x} = -\frac{4}{6}$

What is x^3?

☐ $\frac{-8}{27}$

☐ $\frac{27}{8}$

☐ $\frac{-27}{8}$

☐ $\frac{8}{27}$

Question 3.

How many real solutions does the equation below have?

$(x-2)(x^2-9)=0$

☐ 0

☐ 1

☐ 2

☐ 3

Question 4.

What is the derivative *f'(x)* if

$f(x)=\frac{2}{6x}+\frac{6x}{4}$?

☐ $f'(x)=\frac{1}{3x^2}+\frac{3}{2}$

☐ $f'(x)=\frac{-1}{3}x^{-2}+\frac{3}{2}$

☐ $f'(x)=\frac{2}{6x^{-2}}+\frac{3}{2}$

☐ $f'^{(x)}=-\frac{2}{6x^{-1}}+\frac{6}{4}$

Question 5.

Determine the value for

$\lim_{x\to 0}(e^x+2)$

☐ 0

☐ 2

☐ e (≈ 2.7)

☐ 3

Question 6.

$$\begin{cases} 4y - 2x = 6 \\ 6y - x = 4 \end{cases}$$

What are the values for *y* and *x*?

☐ $y = -0.25$, $x = -2.5$

☐ $y = 0.25$, $x = -2.5$

☐ $y = -0.25$, $x = 2.5$

☐ $y = 0.25$, $x = 2.5$

Question 7.

Which of the following equals 12% of 75?

☐ $\frac{75}{12\times100}$

☐ $\frac{75\times100}{12}$

☐ $\frac{12\times100}{75}$

☐ $\frac{12\times75}{100}$

Question 8.

What equals $(x^4)^3$?

☐ x^7

☐ x^8

☐ x^{12}

☐ x^{64}

Question 9.

What is $\frac{a/b}{c/d}$ if $a = 2c$ and $b = 2d$?

☐ 1

☐ 2

☐ a/b

☐ c/d

Question 10.

What is x if

$3^{-2x+6} = \frac{1}{9}$?

☐ $x = 1$

☐ $x = 2$

☐ $x = 3$

☐ $x = 4$

You scored **higher than __%** of university students in a comparison group who have taken this test before you. This means you performed **better than __%** and **worse or equal to __%** of the students in the comparison group.

In general, how do you **see yourself**? Where would you place yourself on this scale ranging from "Very masculine" = 1 to "Very feminine" = 11? Please indicate your response below.

Very masculine									Very feminine		I prefer not to
1	2	3	4	5	6	7	8	9	10	11	answer
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

In general, how do you **most people** see you? Where would **most people** place you on this scale ranging from "Very masculine" = 1 to "Very feminine" = 11? Please indicate your response below.

Very masculine									Very feminine		I prefer not to answer
1	2	3	4	5	6	7	8	9	10	11	
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

When you finish school and start working, how many **days per month** do you think you would be willing to travel for work? (Travel in this case meaning being away from home over night). Please indicate your response below.[5]

0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30

When you finish university and start working, how many **hours per week** do you think you would be willing to work (if you were compensated accordingly)? Please indicate your response below.

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

How **confident** are you? Where would you place yourself on this scale ranging from "Not confident at all" = 1 to "Very confident" = 11? Please indicate your response below.

Not confident at all									Very confident	
1	2	3	4	5	6	7	8	9	10	11
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

---

[5] Participants indicated their response using a slider, meaning they could choose any number between 1-30.

How **ambitious** are you? Where would you place yourself on this scale ranging from "Not ambitious at all" = 1 to "Very ambitious" = 11? Please indicate your response below.

Not ambitious at all								Very ambitious		
1	2	3	4	5	6	7	8	9	10	11
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

How would you place the subject **mathematics** on this scale ranging from "Very masculine" = 1 to "Very feminine" = 11? Please indicate your response below.

Very masculine								Very feminine		
1	2	3	4	5	6	7	8	9	10	11
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

How would you estimate your **math ability**? Where would you place yourself on this scale ranging from "Very bad at math" = 1 to "Very good at math" = 11? Please indicate your response below.

Very bad at math								Very good at math		
1	2	3	4	5	6	7	8	9	10	11
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Do you remember how many percent (%) of the students in the comparison group you scored better than? Indicate this below.

0 10 20 30 40 50 60 70 80 90 100

What do you think the purpose of this study is?

What is your **legal** gender?

☐ Male

☐ Female

How old are you?

Which program do you study?

☐ Economics (in Swedish ”Nationalekonomi”)

☐ Business Economics/Accounting (in Swedish “Företagsekonomi”)

☐ Civil Economics (in Swedish “Civilekonom”)

☐ Finance (in Swedish “Finans”)

☐ Other program within Economics/Business/Finance

☐ Other program which is **not** Economics/Business/Finance

☐ I am not a university student

At what level are you currently studying? (If you are studying a program which is not a bachelor or master, for example civil economics, please indicate the corresponding year you **would have been on** if your program was a bachelor or master).

☐ Bachelor – Year 1

☐ Bachelor – Year 2

☐ Bachelor – Year 3

☐ Master – Year 1

☐ Master – Year 2

☐ I am not a university student

Appendix B: Additional Figures and Tables

Figures

Stereotype Distribution

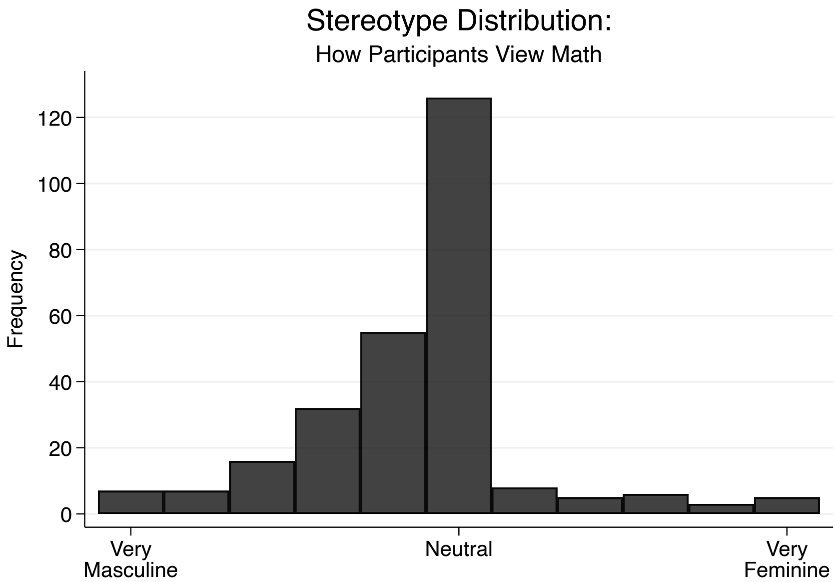

Figure B1. *Stereotype Distribution - How Participants View Math.*

Score Distributions - Reference Groups

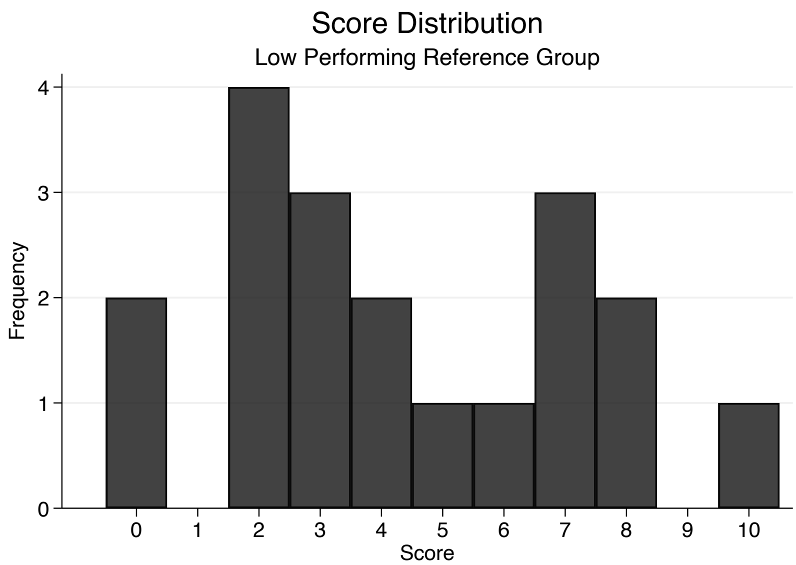

Figure B2. *Score Distribution - Low Reference Group.*

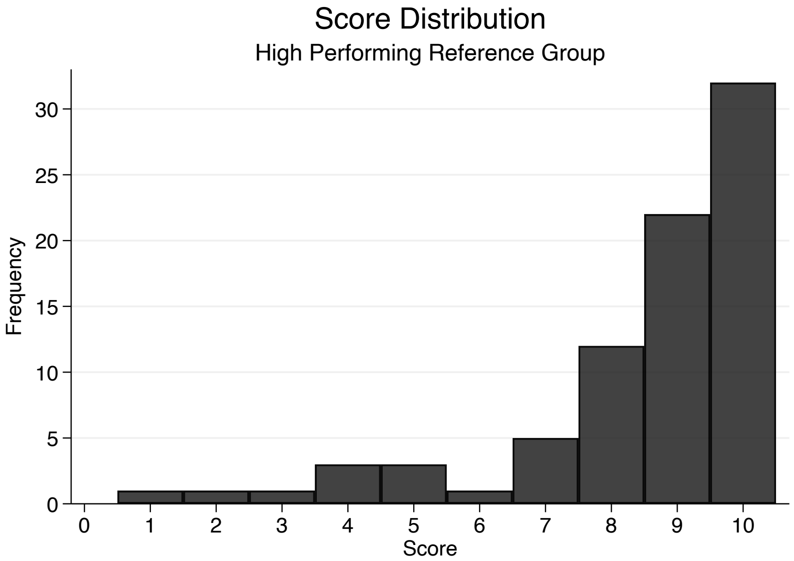

Figure B3. *Score Distribution - High Reference Group.*

Score Distributions - Main Study

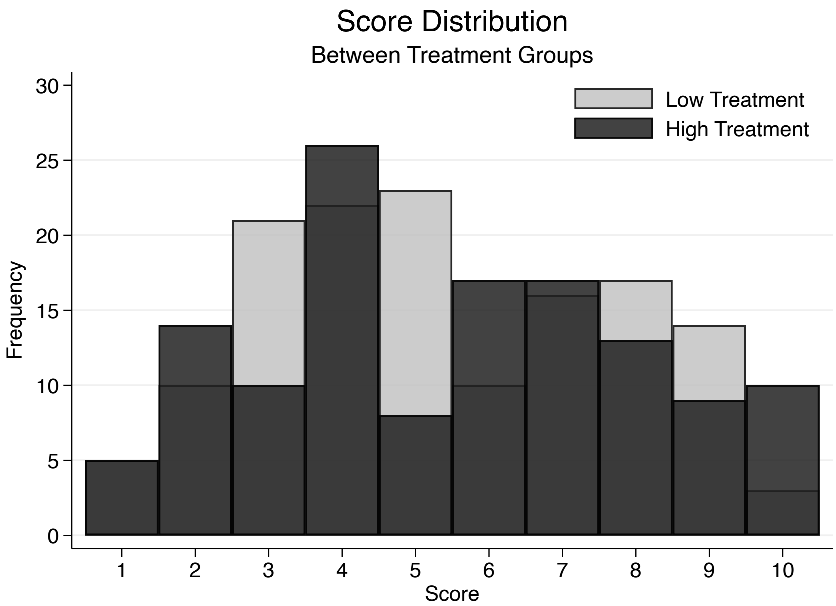

Figure B4. *Score Distribution - Between Treatment Groups.*

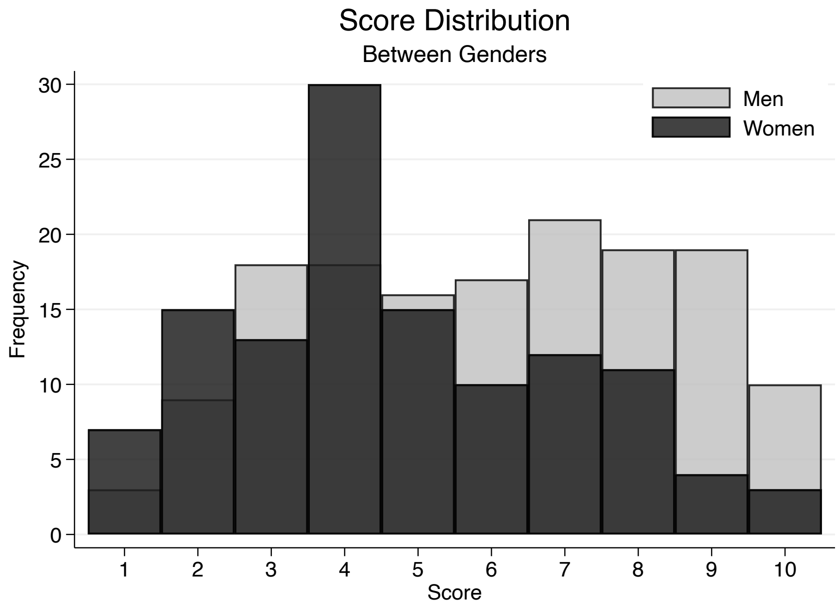

Figure B5. *Score Distribution - Between Men and Women.*

How Participants Believe Others Perceive Their Gender Identity

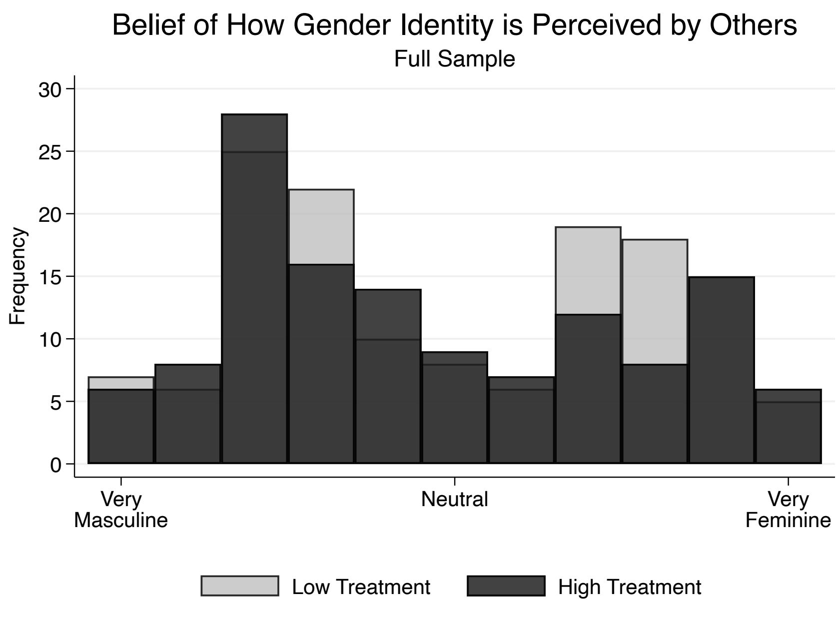

Figure B6. *GI_{Others} Distribution - Full Sample.*

A t-test show no significant difference, $t(268) = 0.919$, $p = 0.359$, between the low ($M = 5.965$, $SD = 2.931$) and high ($M = 5.636$, $SD = 2.941$) treatment groups for the full sample.

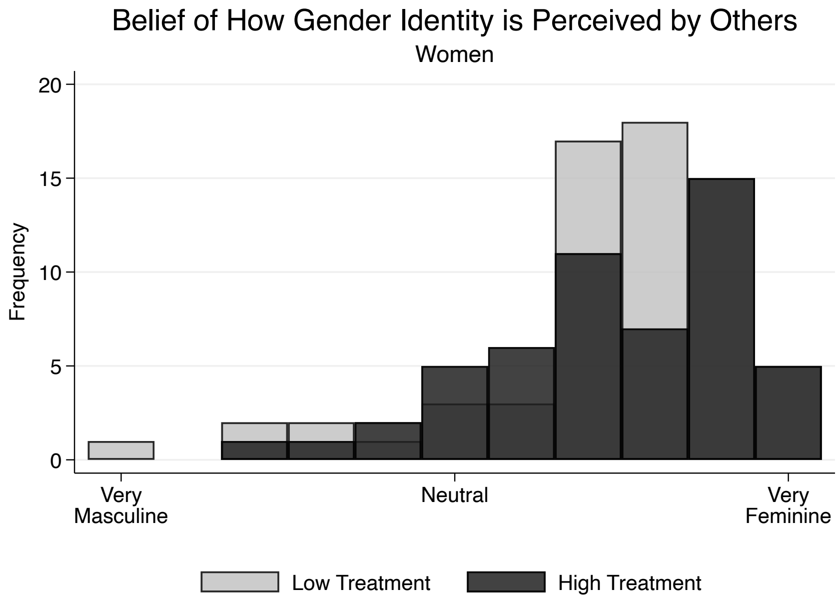

Figure B7. *GI_{Others} Distribution - Women.*

A t-test show no significant difference, *t*(118) = -0.023, *p* = 0.982, between the low (*M* = 8.388, SD = 2.015) and high (*M* = 8.396, *SD* = 1.905) treatment groups for the women in the sample.

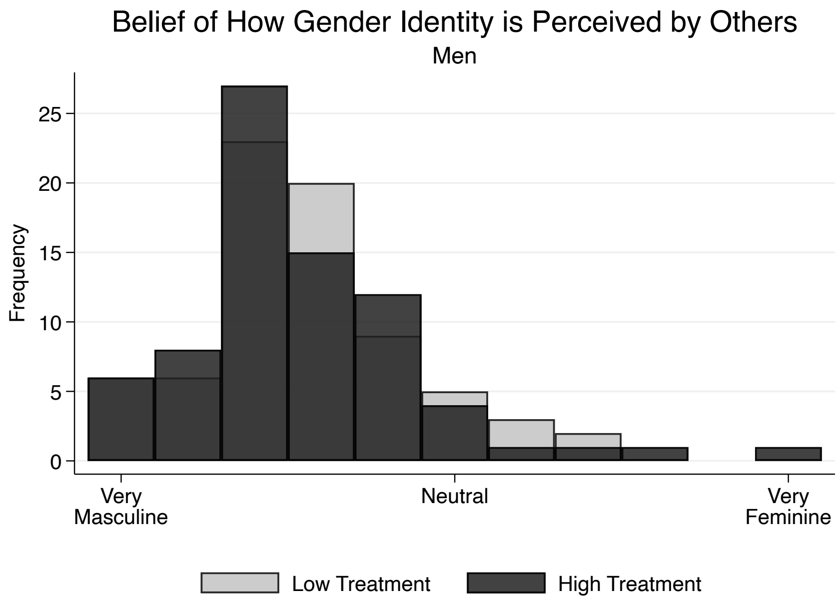

Figure B8. *GI_{Others} Distribution - Men.*

A t-test show no significant difference, *t*(148) = 0.218, *p* = 0.828, between the low (*M* = 3.770, SD = 1.592) and high (*M* = 3.710, *SD* = 1.765) treatment groups for the men in the sample.

Tables

Variables

Table B1.

Variables.

Variable	Description
Gender Identity (*GI*)	Perception of own gender identity. Measured on an 11-point scale where 1 = "Very Masculine" and 11 = "Very Feminine".
Others' Perception of Gender Identity (GI_{Others}):	How participants believe their gender identity is perceived by others. Measured on an 11-point scale where 1 = "Very Masculine" and 11 = "Very Feminine".
High Treatment	Binary variable for treatment where High Treatment = 1 means the participant received feedback where their score was compared to the low performing reference group, thus signaling high relative math performance, and High Treatment = 0 means the participant received feedback where their score was compared to the high performing reference group, thus signaling low relative math performance.
Male	Binary variable for legal gender where 1 = "Male" and 0 = "Female".
Ability	Participants' perception of their math ability. Measured on an 11-point scale where 1 = "Very bad at math" and 11 = "Very good at math".
Willingness to Travel	How many days per month the participant would be willing to travel for work. Measured on a scale ranging between 0-30 days per month.

Hours Willing to Work	How many hours per week the participant would be willing to work. Measured on a scale ranging between 0-70 hours per week.
Confidence	Participants' perception of their confidence. Measured on an 11-point scale where 1 = "Not confident at all" and 11 = "Very confident".
Ambition	Participants' perception of their ambition. Measured on an 11-point scale where 1 = "Not ambitious at all" and 11 = "Very ambitions".
Score	How many correct answers the participant had on the math test. Value between 0-10.
Program	Categorical variable for program where 1 = "Business", 2 = "Economics", 3 = "Business & Economics", 4 = "Finance", 5 = "Other Related Program", 6 = "Other Unrelated Program".

Manipulation Check

Table B2.

Treatment on Estimated Math Ability.

	Full Sample	Men	Women
	(1)	(2)	(3)
High Treatment	0.514* (0.270)	0.650* (0.364)	0.312 (0.407)
Constant	6.277*** (0.192)	6.297*** (0.268)	6.254*** (0.275)
N	270	150	120
R^2	0.013	0.021	0.005

Notes: Dependent variable is estimated math ability, which is measured on a scale from 1 = "Very Bad at Math" to 11 = "Very Good at Math". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Robustness Checks

Remembered Feedback

Table B3.

Treatment on GI: Remembered Feedback.

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.365 (0.420)	0.024 (0.322)	-0.130 (0.392)	-0.839** (0.424)
Constant	5.824*** (0.287)	3.559*** (0.206)	8.551*** (0.246)	5.792*** (0.296)
N	206	119	87	187
R^2	0.004	0.000	0.001	0.021

Notes: Only participants who could remember their performance feedback included in sample. Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B4.

Treatment on GI_{Others}: Remembered Feedback.

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.319 (0.412)	0.057 (0.316)	-0.092 (0.404)	-0.872** (0.414)
Constant	5.778*** (0.286)	3.593*** (0.200)	8.408*** (0.281)	5.802*** (0.297)
N	206	119	87	187
R^2	0.003	0.000	0.001	0.023

Notes: Only participants who could remember their performance feedback included in sample. Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B5.

Treatment on GI – Gender Differences: Remembered Feedback.

	Full Sample	Excl. Feminine Stereotypes
	(1)	(2)
High Treatment	-0.130 (0.391)	-0.498 (0.422)
Male	-4.992*** (0.320)	-5.040*** (0.325)
High Treatment × Male	0.154 (0.507)	0.235 (0.514)
Constant	8.551*** (0.246)	8.636*** (0.250)
N	206	187
R^2	0.657	0.695

Notes: Only participants who could remember their performance feedback included in sample. Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B6.

Treatment on GI_{Others} – Gender Differences: Remembered Feedback.

	Full Sample	Excl. Feminine Stereotypes
	(1)	(2)
High Treatment	-0.092 (0.403)	-0.717* (0.429)
Male	-4.815*** (0.345)	-4.982*** (0.342)
High Treatment × Male	0.149 (0.513)	0.507 (0.517)
Constant	8.408*** (0.281)	8.614*** (0.276)
N	206	187
R^2	0.634	0.675

Notes: Only participants who could remember their performance feedback included in sample. Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Understood the Purpose of the Study

Table B7.

Treatment on GI: Excluding Those Who Understood the Purpose.

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.352 (0.372)	-0.099 (0.290)	0.164 (0.322)	-0.815** (0.379)
Constant	6.037*** (0.254)	3.667*** (0.204)	8.478*** (0.212)	6.04*** (0.262)
N	263	143	120	236
R^2	0.003	0.001	0.002	0.019

Notes: Only people who did not understand the purpose of the study included in sample. Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B8.

Treatment on GI_{Others}: Excluding Those Who Understood the Purpose.

	Full Sample	Men	Women	Excl. Feminine Stereotypes
	(1)	(2)	(3)	(4)
High Treatment	-0.411 (0.361)	-0.125 (0.284)	0.008 (0.359)	-0.906** (0.369)
Constant	6.088*** (0.250)	3.855*** (0.194)	8.388*** (0.246)	6.104*** (0.262)
N	263	143	120	236
R^2	0.005	0.001	0.000	0.025

Notes: Only people who did not understand the purpose of the study included in sample. Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B9.

Treatment on GI – Gender Differences: Excluding Those Who Understood the Purpose.

	Full Sample (1)	Excl. Feminine Stereotypes (2)
High Treatment	0.164 (0.322)	-0.196 (0.334)
Male	-4.811*** (0.294)	-4.868*** (0.289)
High Treatment × Male	-0.263 (0.433)	-0.190 (0.429)
Constant	8.478*** (0.212)	8.610*** (0.204)
N	263	236
R^2	0.670	0.712

Notes: Only people who did not understand the purpose of the study included in sample. Dependent variable is Gender Identity (GI), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B10.

Treatment on GI_{Others} – Gender Differences: Excluding Those Who Understood the Purpose.

	Full Sample (1)	Excl. Feminine Stereotypes (2)
High Treatment	0.008 (0.359)	-0.552 (0.387)
Male	-4.533*** (0.313)	-4.682*** (0.318)
High Treatment × Male	-0.134 (0.458)	0.201 (0.471)
Constant	8.388*** (0.246)	8.576*** (0.249)
N	263	236
R^2	0.615	0.647

Notes: Only people who did not understand the purpose of the study included in sample. Dependent variable is Others' Perception of Gender Identity (GI_{Others}), which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Excluding Those with Feminine or No Stereotypes

Table B11.

Treatment on GI & GI_{Others}: Only Masculine Stereotypes.

	GI	GI_{Others}
	Excl. Feminine or No Stereotypes	Excl. Feminine or No Stereotypes
	(1)	(2)
High Treatment	-0.943* (0.564)	-0.869* (0.519)
Constant	5.860*** (0.405)	5.719*** (0.393)
N	117	117
R^2	0.024	0.024

Notes: Only participants with strictly masculine stereotypes included. Dependent variable is Gender Identity (GI) in column 1 and Others' Perception of Gender Identity (GI_{Others}) in column 2, which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.

Table B12.

Treatment on GI & GI_{Others} – Gender Differences: Only Masculine Stereotypes.

	GI	GI_{Others}
	Excl. Feminine or No Stereotypes	Excl. Feminine or No Stereotypes
	(1)	(2)
High Treatment	-0.204 (0.521)	-0.613 (0.641)
Male	-5.024*** (0.461)	-4.561*** (0.533)
High Treatment × Male	-0.452 (0.648)	0.228 (0.729)
Constant	8.68*** (0.329)	8.28*** (0.455)
N	117	117
R^2	0.722	0.618

Notes: Only participants with strictly masculine stereotypes included. Dependent variable is Gender Identity (GI) in column 1 and Others' Perception of Gender Identity (GI_{Others}) in column 2, which is measured on a scale from 1 = "Very Masculine" to 11 = "Very Feminine". High treatment indicates receiving feedback of high relative math performance, i.e., being compared to the low performing reference group. Robust standard errors in parentheses. ***, ** and * indicate significance at 1%, 5% and 10% levels.