



University of
Zurich^{UZH}

URPP Equality of Opportunity

Gender-Stereotypes in Task Choice: An Experimental Investigation of Taste-Based vs. Statistical Self- Stereotyping

Lea Naemi Weigand

Equality of Opportunity Research Series #84
December 2025





**University of
Zurich** ^{UZH}

URPP Equality of Opportunity

URPP Equality of Opportunity Discussion Paper Series No. 84, December 2025

Gender-Stereotypes in Task Choice: An Experimental Investigation of Taste-Based vs. Statistical Self-Stereotyping

Lea Naemi Weigand

Gutenberg University Mainz and Goethe University Frankfurt

weigandl@uni-mainz.de

The University Research Priority Program “Equality of Opportunity” studies economic and social changes that lead to inequality in society, the consequences of such inequalities, and public policies that foster greater equality of opportunity. We combine the expertise of researchers based at the University of Zurich’s Faculty of Arts and Social Sciences, the Faculty of Business, Economics and Informatics, and the Faculty of Law.

Any opinions expressed in this paper are those of the author(s) and not those of the URPP. Research published in this series may include views on policy, but URPP takes no institutional policy positions.

URPP Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

URPP Equality of Opportunity, University of Zurich, Schoenberggasse 1, 8001 Zurich, Switzerland
info@equality.uzh.ch, www.urpp-equality.uzh.ch

Gender-Stereotypes in Task Choice: An Experimental Investigation of Taste-Based vs. Statistical Self-Stereotyping

Lea Naemi Weigand*

December 23, 2025

Abstract

Self-stereotyping describes the behavior of applying stereotypes to oneself. Using a novel design in an online experiment, I analyse two types of self-stereotypical behavior: statistical and taste-based self-stereotyping. Statistical self-stereotyping is defined as self-stereotypical behavior driven by stereotyped beliefs about one's own abilities. Taste-based self-stereotyping refers to self-stereotyping due to intrinsic preferences for stereotype compliance. In the experiment, the stereotype participants associate with a given task is exogenously changed via differential task framing holding all other task characteristics fixed. Exploiting this exogenous shift in task stereotype, I find statistical self-stereotyping among women: Framing the same task as a male-typed task instead of a female-typed task has a significant and negative impact on women's beliefs about their future task performance, conditional on past performance. I find no strong evidence for taste-based self-stereotyping. The findings highlight the subtle yet powerful influence of task stereotypes on beliefs about own ability and underscore the importance of stereotype context in shaping individual confidence.

*Department of Economics, Gutenberg University Mainz and Goethe University Frankfurt; weigandl@uni-mainz.de. I am grateful to Roberto Weber for his guidance and support in the development of this project. I thank Sandro Ambühl, Kai Barron, Anne Brenøe, Katie Coffman, René Karadacic, Claudia Marangon, Raphaël Raux, Simona Sartor, and Jeffrey Yusof for helpful comments, feedback and discussions. I gratefully acknowledge financial support from the University of Zurich *URPP Equality of Opportunity*.

1 Introduction

In many contexts, individuals apply stereotypes to themselves, a phenomenon referred to as self-stereotyping. Gender-stereotypes in particular have been shown to matter for a variety of behaviors importantly related to gender differences in labor market outcomes such as willingness to compete (Kamas and Preston, 2012; Dreber et al., 2014; Grosse et al., 2014), to lead (Chen and Houser, 2019; Alan et al., 2020), and to express ambition (Bursztyn et al. 2017). Given these interrelations, it is crucial to understand the channels through which self-stereotyping occurs, not only in its own right, but to successfully tackle related labor market gender differences.

With this project, I contribute to our understanding of self-stereotyping. To do so, I run an online experiment to shed light on two distinct potential drivers of self-stereotypical behavior in the same setting: *statistical self-stereotyping* and *taste-based self-stereotyping*. In addition, I introduce a novel experimental design to study self-stereotyping generally. This design varies the gender stereotype attached to an otherwise identical task via differential task framing to isolate the impact of the task stereotype on outcomes of interest.

Analogous to statistical discrimination, statistical self-stereotyping refers to individuals using stereotypes about a group to which they belong to draw conclusions about themselves, often regarding their own abilities. This can lead to biased beliefs about one's own skills. Taste-based self-stereotyping instead, similar to the concept of taste-based discrimination, describes individuals' intrinsic preferences to self-stereotype, i.e., to act in accordance with stereotypes.

In the experiment in this paper, participants work on a matrix task. They are randomly assigned to the task being presented as a male- or as a female-typed the task. Subjects are then asked to work on the task. To measure statistical self-stereotyping, I elicit subjects' beliefs about their ability in the task after providing feedback on their previous task performance. I find statistical self-stereotyping for women: Conditional on task performance, women assigned to the male-framed version of the task believe they will perform significantly worse in the task in the future compared to women assigned to the female version of the task. Then, I turn to the analysis of taste-based self-stereotyping. I implement a revealed preference measure for the matrix task which elicits participants' choice of a task to perform in a second round of task work. Taste-based self-stereotyping predicts that participants choose to perform the matrix task relatively more often (hence, expressing a preference

for this task) if assigned to the gender-congruent framing compared to the gender-incongruent framing. I do not find evidence confirming this hypothesis in my study. To better understand the observed behavior, I conduct heterogeneity analyses by baseline task performance, the direction of belief-updating after feedback and whether the task framing worked as intended for the respective individual.

The concept of statistical self-stereotyping has been formalized by [Bordalo et al. \(2016\)](#) who set-up a model which predicts that gender stereotypes influence the assessment of one's own abilities. This can then discourage counter-stereotypical behaviors such as the participation in gender-incongruent occupations: Individuals believe that their gender, on average, performs worse than opposite-gender peers in such jobs and conclude that they themselves will perform less well than their opposite-gender colleagues. Even if such beliefs about a group's average job performance were accurate, they only describe a summary statistic for the respective group and hence never perfectly predict an individual's ability in a given occupation. For this reason, statistical self-stereotyping overall leads to suboptimal choices. In the context of occupation choices, for example, this results in an inefficient allocation of talent in the aggregate economy.

Current policy interventions to tackle self-stereotyping mostly focus on counteracting statistical self-stereotypical by correcting beliefs about individuals' abilities in counter-stereotypical domains. This makes sense given that previous research shows that stereotyped beliefs about one's abilities are an important driver of self-stereotypical behavior (e.g., [Kamas and Preston 2012](#); [Coffman 2014](#); [Dreber et al. 2014](#)). [Coffman \(2014\)](#), for instance, finds that one channel through which stereotypes affect contributions to a group outcome across male- and female-typed tasks are beliefs about ability. To counteract these beliefs, [Coffman \(2014\)](#) implements an intervention which provides individuals with information on their strengths in a counter-stereotypical task. This aims at correcting misperceptions about ability and reducing the scope for self-stereotyping. However, [Coffman \(2014\)](#) finds that their intervention is not successful in promoting individuals' contributions to the group outcome in counter-stereotypical task domains. This suggests that the intervention was either unsuccessful in affecting beliefs or that beliefs are not the only force driving the observed self-stereotypical behavior in this setting. If the latter is indeed the case, correcting misperceptions about abilities is not sufficient to successfully counteract self-stereotypical behavior.

One plausible alternative channel through which self-stereotyping could occur is through prefer-

ences for stereotypical behavior. I define taste-based self-stereotyping as self-stereotypical behavior that is determined by an individual's inherent preference or distaste for complying with existing stereotypes, independent of beliefs about their own abilities. Going back to the example of occupation choices, this captures any hesitancy of individuals to apply to gender-incongruent jobs due to their own preference for stereotype-compliance in their occupation choice. So far, there is only suggestive evidence that intrinsic preferences for stereotype-compliance exist. [Coffman et al. \(2021a\)](#) analyse how teams decide which group member should answer a quiz question on behalf of the group. They find that this decision is done in a way that is consistent with gender-stereotypes (men are chosen to answer questions in male-typed quiz domains, women are chosen to answer questions in female-typed domains) but cannot be fully rationalized by beliefs about the abilities of the chosen group members. The authors conclude that inherent preferences to act in accordance with stereotypes might play a role in explaining their findings ([Coffman et al. 2021a](#)). However, their setting does not allow to draw causal conclusions on this hypothesis. Further, it is hard to disentangle the impact of self-stereotyping and stereotyping by others in their context. In a different but related setting, [Exley and Kessler \(2022\)](#) focus on self-promotion in a male-typed task related to math and science. In the treatment condition most relevant to my study, participants evaluate how likely it is that they would apply for a follow-up job based on their past performance in the experimental task. Even though subjects learn their absolute and relative task performance prior to the decision to self-promote, the authors find a significant gender gap in self-promotion: In the male-typed task considered in the study, women self-promote significantly less than men, conditional on past performance. This could be driven both by statistical as well as taste-based self-stereotyping: If the latter is the case, female (male) participants have a distaste (preference) for working in a male-typed job which translates into a decreased (increased) likelihood of applying conditional on ability. The setting in [Exley and Kessler \(2022\)](#) does not allow to test for this hypothesis directly. An alternative explanation would be statistical self-stereotyping: If beliefs about general task ability after having received feedback on past performance depend positively on the gender-congruency of the task, women end up holding less optimistic beliefs about their task ability in the male-typed task compared to men, even conditional on past performance. Such behavior has been documented in [Coffman et al. \(2021b\)](#) and could also explain the gender gap in self-promotion observed in [Exley and Kessler \(2022\)](#). In this project, I aim at measuring both mechanisms plausibly underlying

self-stereotyping.

The rest of this paper is structured as follows. Section 2 describes the experimental design. Section 3 presents the main analyses and corresponding results. Section 4 concludes.

2 Experimental Design

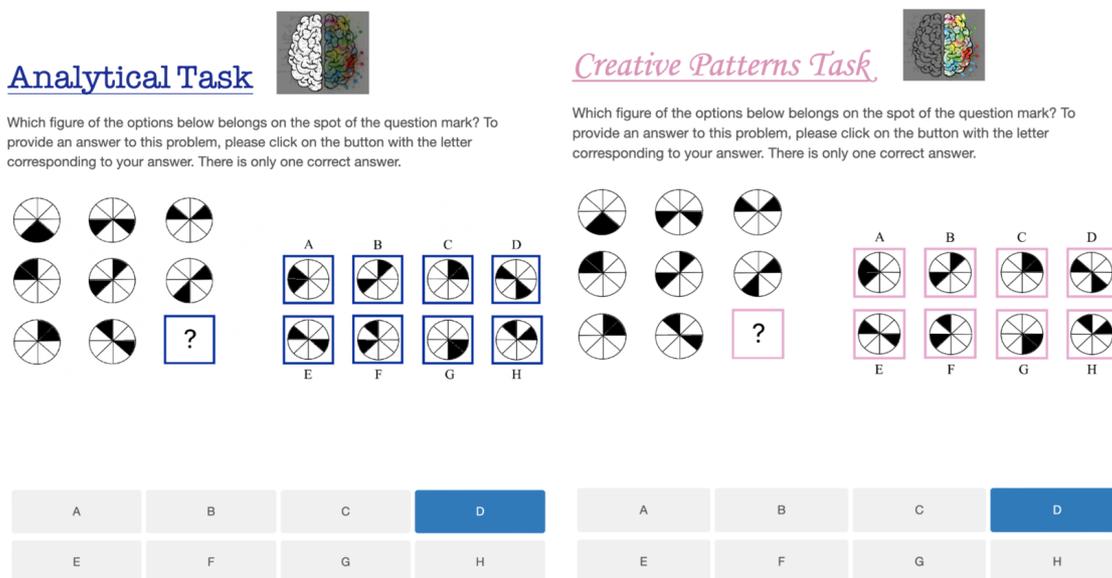
In this project, the main goal of the experiment is to test for the two channels of self-stereotypical behavior previously described: statistical and taste-based self-stereotyping. To do so, I vary the gender stereotype attached to the experimental task using differential task framing. This allows me to isolate the impact of task stereotype on beliefs about task ability (statistical self-stereotyping) and preferences for the task (taste-based self-stereotyping). The experiment was pre-registered at the [AEA RCT Registry](#) with RCT ID: AEARCTR-0013730.

2.1 Main Idea: Inducing Variation in Task Stereotype using Framings

Experiments on self-stereotyping usually compare the outcome of interest across settings that vary in their gender-stereotype. For example, [Kamas and Preston \(2012\)](#) compare competitive behavior of women and men in verbal (female-typed) versus math (male-typed) tasks. In [Coffman \(2014\)](#) and [Coffman et al. \(2021a\)](#) participants take part in quizzes across various domains that differ in their gender-stereotype. Hence, in these studies the variation in the gender-stereotype is exogenously induced by changing either the task itself or the task domain. While this is clearly effective in changing the task stereotype, it can cause omitted variable bias if the change in the task (domain) not only shifts the stereotype but also other task characteristics that are correlated with both the stereotype and the outcome of interest. One example for such an omitted variable is task experience: If individuals have more experience in the stereotypical task (domain), differential behavior across tasks could be driven by task experience and not by the stereotype attached to the task. To avoid this problem of omitted variables in my experiment, I change the task stereotype while holding the task completely fixed. I do so by framing the same task as a either male- or female-typed task. The main idea for the experiment is based on [Barron et al. \(2023\)](#) who investigate the impact of the gender-stereotype associated with a task on the competitive behavior of men and women. However, the concrete experimental design in my experiment is adjusted compared to [Barron et al. \(2023\)](#) to

incorporate insights from pre-tests which were run to optimally design the framings to successfully shift the perceived task stereotype.

Figure 1: Task framings



Notes: This figure presents the task framings used in the study to present the matrix task as a male-typed task (left) or a female-typed task (right).

As an illustration of the main idea of changing the task stereotype through framing, consider a matrix task as displayed in Figure 1. In the task, study participants have to logically complete the pattern on the left with one of the options given on the right. While the task itself remains unchanged, the presentation on the left of Figure 1 presents the task as a male-typed “Analytical Task” and the version on the right displays the same task as a female-typed “Creative Patterns Task”. Both framings appear credible given the nature of the task as a test of logical thinking on the one hand, and the involvement of shapes and patterns on the other. The framing manipulation hence changes the assigned task stereotype while keeping the task completely the same in all substantive regards.

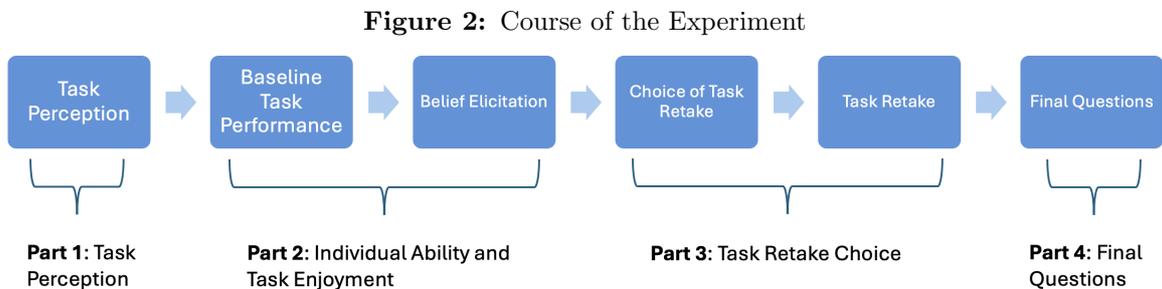
In the experiment, the treatment consists of the random assignment of experimental subjects to either the male or the female task framing. After being assigned to one of the two versions of the task, subjects in the experiment will work on the respective task and answer questions on their beliefs about their task ability, task enjoyment and a revealed preference measure for the task. The exogenous variation in the task stereotype induced by the task framings allows me to investigate

the impact of the stereotype on these outcomes to analyse statistical self-stereotyping (via the impact on beliefs) and taste-based self-stereotyping (via the impact on enjoyment and revealed task preference).

I have pre-tested these framings to make sure they successfully shift the perceived task stereotype. They do: Survey participants think that most people evaluate the task as more feminine (masculine), believe that women (men) perform better in the task and believe that women (men) enjoy working on the task more if the task is female-framed (male-framed). Hence, the framings work in the intended way and provide a novel way to exogenously induce a change in the task stereotype within an experimental setting.¹

2.2 Course of the experiment

Figure 2 gives a visual overview of the experiment which will be described in detail below. Before the main part of the experiment starts, I elicit consent as well as gender and age of the participants. In the experiment, I first randomly allocate every participant to one of the two framings described before. The experiment is designed as a between-subject study, hence every participant will only see one of the two framings throughout the experiment.



Notes: This figure gives a broad overview of the different parts of the experiment.

2.2.1 Part 1: Task Perception

In part 1 of the experiment, I measure the degree to which participants perceive their assigned (framed) task as stereotypically male or female. Importantly, each study subject evaluates the task only with the framing they are randomly assigned to. Eliciting task perception is important to

¹For more detailed results from pre-testing please see Appendix Section A. The instructions for eliciting task perceptions in the pre-tests were the same as in the main survey.

establish that the different framings actually shift the perceived task gender-stereotype. I have pre-tested the framings to make sure they work as intended but I confirm the successful shift in task stereotype in my main sample as well. To do so, I follow [Krupka and Weber \(2013\)](#) and elicit participants' "coordinated beliefs" about the task. I use the term "coordinated beliefs" to describe that participants are incentivized to match with their task evaluation the evaluation most frequently provided by other study subjects. I elicit three dimensions of task perception to capture various concepts of stereotypes used in the existing literature.

Whether participants think most people believe ...

- ... the task is masculine or feminine.
- ... men or women perform better in the task.
- ... men or women enjoy working on the task more.

I elicit respondents' "coordinated beliefs" for two main reasons: First, these beliefs seem to be the relevant metric in the context of stereotypes. Stereotypes exist only as a construct of society in which individuals agree to label certain things and behaviors as stereotypically male or female. Hence, when evaluating an activity in terms of being stereotypical we essentially ask about the view society has adopted on the gender-congruency of the corresponding behavior. In addition, measuring "coordinated beliefs" allows to incentivize answers.²

2.2.2 Part 2: Individual Ability and Task Enjoyment

After eliciting task perceptions, experimental subjects work on the assigned task for one round. A round consists of 8 different matrix problems. I refer to this first round of work as the baseline work round. Participants have a fixed amount of time of 25 seconds to work on each problem. They cannot continue to the next problem before the 25 seconds are up and are automatically directed to the next problem once the 25 seconds are over. I enforce this timing to make sure that all subjects spend the same amount of time on each problem and do not rush through the task to finish more quickly. In addition, it is important for part 3 of the experiment to know exactly how much time

²In my study, experimental subjects are informed that in each of the three task evaluation dimensions outlined before a subsample is randomly drawn with a quota of 1 in 25. Individuals in the subsample will then receive a bonus payment of USD 0.5 if the answer they provide in the task evaluation corresponds to the answer most frequently chosen by other study participants.

participants spend on completing one round of the matrix task. Task completion is incentivized by performance-based pay of USD 0.20 for each correctly solved problem.

After finishing the baseline work round, subjects are asked how many problems they believe they have solved correctly and how much they enjoyed working on the task (on a 10-point Likert scale). In the experiment, statistical self-stereotyping is captured via beliefs about task performance while taste-based self-stereotyping (at least to some degree) can be reflected in reported task enjoyment levels. Hence, eliciting this information allows me to test two pre-registered hypotheses regarding statistical and taste-based self-stereotyping (Hypotheses 1.1 and 2.1, respectively).

***Hypothesis 1.1** Prior to feedback, study participants believe they have performed better in the matrix task in the baseline work round if the task is stereotypically framed compared to counter-stereotypically, conditional on baseline performance.*

***Hypothesis 2.1** Conditional on beliefs about task ability and baseline performance, study participants report to enjoy working on the matrix task more if the task is stereotypically framed compared to counterstereotypically.*

Next, participants receive perfect feedback on their baseline performance: They learn exactly how many matrix problems they have solved correctly in the baseline work round. I make sure participants pay attention to the feedback provided by asking them to report back the number of matrix problems they have solved correctly at baseline.

After the provision of feedback, I elicit participants' beliefs about their future performance in the task: If they were to redo the task with new problems³, how many problems do subjects expect to solve correctly? Given the perfect feedback on baseline task performance, subjects can form well-founded beliefs about their ability in the task. To encourage truth-telling, I incentivize belief-reporting by providing a bonus payment of USD 0.15 to participants if beliefs match future task performance. In addition, I measure individuals' expected task enjoyment if they were to work on the task for a second time (again, on a 10-point Likert scale). As before, the elicited information is directly relevant for the next set of pre-registered hypotheses on statistical and taste-based self-stereotyping (Hypotheses 1.2 and 2.2, respectively):

***Hypothesis 1.2** Conditional on performance at baseline and feedback, study participants believe*

³Subjects are informed that these new problems are similar in style and difficulty to the ones they have already seen at baseline.

they will perform better in the matrix task in work round 2 if the task is stereotypically framed compared to counterstereotypically.

Hypothesis 2.2 *Conditional on beliefs about task ability and baseline performance, study participants report to expect to enjoy working on the matrix task in work round 2 more if the task is stereotypically framed compared to counterstereotypically.*

There are several plausible ways to relate Hypotheses 1.1 and 2.1 to Hypotheses 1.2 and 2.2. One possibility is that the correlations between task stereotype and beliefs about task performance and enjoyment are stronger in the baseline round than the correlations between task stereotype and beliefs about future performance and enjoyment in work round 2. This could be the case if real experiences during baseline task performance, independent of feedback, play a crucial role in shaping these correlations. Alternatively, it is also plausible that expectations about future performance and enjoyment are more strongly correlated with task stereotype than beliefs about baseline performance. Such a pattern could arise if correlations are driven less by actual baseline experiences and more by "hypothetical" thinking about an uncertain future, plausibly informed by performance feedback.

2.2.3 Part 3: Task Retake Choice

In the third part of the study, participants choose the task they will work on in a second work round. Subjects can choose between the matrix task they have already performed in the baseline work round (framed in the same way as at baseline) and an outside option task. The outside option task is designed to be gender-neutral. Here, participants are asked to click on a specific letter on a screen showing the alphabet. The letter to click on as well as the order of the alphabet changes every 20 seconds. Participants have to work on this task for 3 minutes and 20 seconds. This fixed amount of time is exactly the same as the time subjects would spend on the second round of the matrix task (where they again face 8 problems and are forced to spend 25 seconds on each problem). If subjects choose the outside option task, they receive a fixed payment for it. To elicit which of the two tasks participants want to complete in work round 2, I use a multiple price list (MPL) in which subjects have to decide between the matrix and the outside option task across 9 decision rows (see figure 3 showing the MPL for subjects assigned to the female-typed framing).

Figure 3: Multiple Price List

Part 2

In Part 2 of the study, you will again perform a task. You can **choose** if you would like to **complete the Creative Patterns Task again** or if you want to perform a Clicking Task.

Please make a choice in each of the 9 rows:

<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$0.00
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$0.20
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$0.40
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$0.60
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$0.80
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$1.00
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$1.20
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$1.40
<u>Creative Patterns Task</u> for \$0.20 per correctly solved problem	<input type="radio"/> <input type="radio"/>	<u>Clicking Task</u> for a total payment of \$1.60

Notes: This figure shows the multiple price list (MPL) used to elicit a revealed preference measure for the matrix task. The MPL depicted here is shown to participants assigned to the female-typed framing, while participants assigned to the male-typed framing see the choice “Analytical Task” on the left side of the MPL.

Across these rows, the piece-rate earned for each correctly solved matrix problem is held constant at USD 0.20 (the same rate as in the baseline round) while the fixed payment earned for performing the outside option task is given by all USD 0.20 increments between USD 0.00 and USD 1.60. These amounts are chosen to match all potential earnings a participant can earn in the matrix task. After filling out the MPL, one row is randomly implemented to determine the task a subject will work on in work round 2. Hence, all decisions in the multiple price list could potentially determine the task in the second work round. This is explicitly pointed out to subjects to incentivize decisions that reflect subjects’ true preferences across the entire price list. The MPL presents a revealed preference

measure for the matrix task providing an additional way to test taste-based self-stereotyping without relying on self-reported task enjoyment. More precisely, I test the following hypothesis:

***Hypothesis 2.3** Conditional on beliefs about task ability, study participants choose to perform the matrix task in work round 2 relatively more often if the task is stereotypically framed compared to counterstereotypically.*

2.2.4 Part 4: Demographics and Final Questions

In the final part of the experiment, I measure character traits (risk aversion, self-confidence, social desirability, importance of being popular, the degree of fitting in, challenge seeking, a measure of continuous gender and Big Five) and demographic information (education, field of study/occupational industry and the number of siblings). I also measure gender norms by eliciting participants' level of agreement to a statement on equally sharing household and market work between a man and a woman within a household. Lastly, I elicit subjects' perceptions of the difficulty of the matrix task and what they believe the matrix task is measuring.

2.3 Sample

I run the experiment via the online survey platform [Qualtrics](#). Subjects are recruited through [Prolific](#), a recruitment platform for participants for online studies. I recruit 1400 participants from the US aged between 18 and 25 years. I focus on this rather homogeneous group of individuals to avoid obscuring treatment effects through opposing effects for different demographic groups. It is a priori unclear for which age group we expect self-stereotyping based on gender to be particularly prevalent but given the many consequential life decisions individuals take in the early years of adulthood (e.g., education decisions and occupation choices), I believe that understanding self-stereotypical behaviour better is especially important for individuals in their formative years. In particular, suggestive evidence on the importance of social norms and stereotypes in the context of education and occupation choices have been documented in a variety of studies (e.g., [Brenøe and Zölitz 2020](#), [Zölitz and Feld 2021](#), [Brenøe 2022](#)). Improving our understanding of self-stereotypical behavior, hence, seems particularly relevant for this group. I restrict my sample to individuals from the US mostly because the pool of available study participants on Prolific is particularly large for

this nationality. In addition, many countries resemble the US in terms of being western liberal democracies for which my findings are thus especially relevant. I recruit an equal number of men and women (700 men and 700 women) and explicitly do not include transgender participants in this sample as I am expecting gender stereotypes to affect cis-gender individuals most.

Prior to implementation, I expected the study to take around 25 minutes to complete. In my sample, 82% of participants completed the study in 25 minutes or less. I pay participants a fixed compensation of USD that amounts to an hourly wage of USD 12. All participants can earn additional bonus payments of up to USD 4.85 throughout the study.

3 Results

As pre-registered, comparisons between treatment conditions in this project are done within gender as it depends on the gender of a study participant whether a task framing is gender-stereotypical or counter-stereotypical. In addition, it is a priori unclear if men and women react in the same way to the (counter-)stereotypical tasks. Hence, pooling the analyses for men and women, while potentially interesting, does not seem to make sense *ex ante*.

3.1 Descriptive Statistics

3.1.1 Demographics

Table 1 describes the sample in detail. In total, I recruited 1,400 individuals out of which 4 had to be dropped as they provided contradicting information on their demographics (gender and age) on Prolific and Qualtrics. This leads to a final sample size of 1,396. The treatment was assigned randomly such that approximately 50% of the sample was assigned to the each framing. The last column of Table 1 shows p-values of mean-comparison tests of demographics for samples by treatment assignment (i.e., framing), providing evidence on the balancedness of treatment groups. In terms of gender distribution, the p-value of 0.914 indicates no significant difference between treatment conditions. Participants in the study are aged between 18 and 25 years. Individuals above 20 years of age are represented in larger numbers than individuals aged 20 and younger. This makes sense given that the pool of available subjects on Prolific. While there are no significant differences in the mean age between both treatment groups, we see some differences across the age

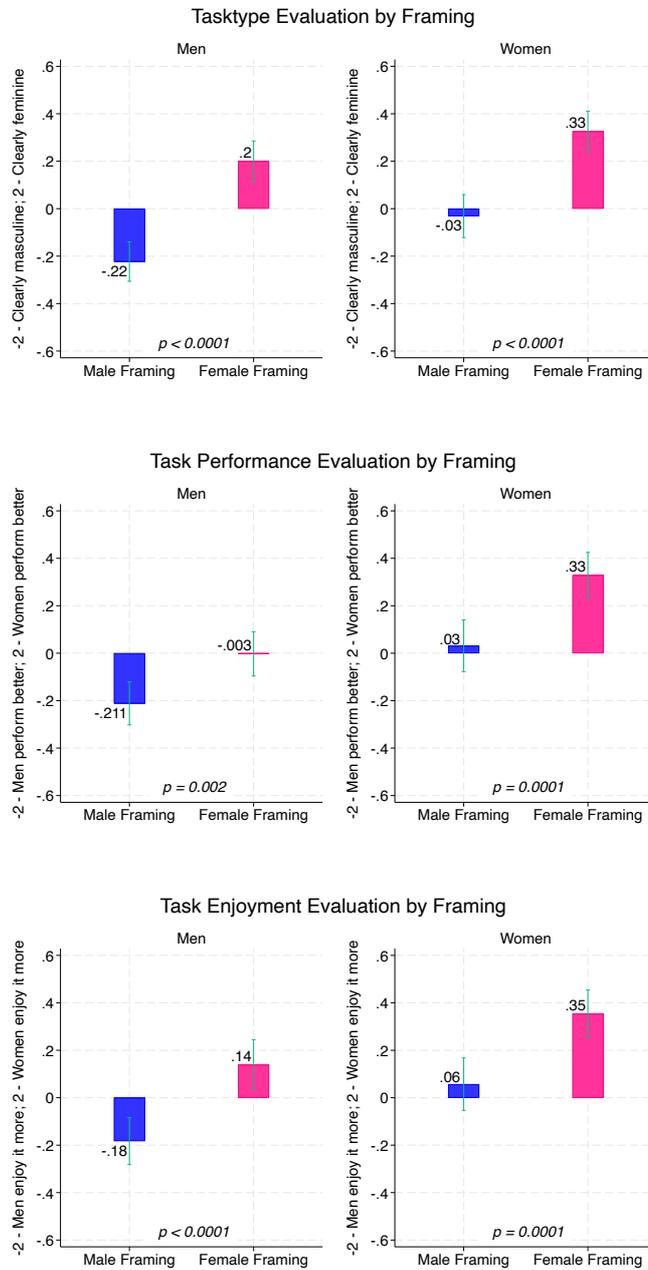
distribution. In the main analyses, I control for age to mitigate concerns regarding these differences. In terms of education, most participants are currently in college (29.7%) or have already obtained a bachelors degree (39%). The majority of participants in both treatment groups is white. Overall, there are no significant differences in the distribution of education and ethnicity across treatment conditions.

Table 1: Descriptives by Treatment Status

	Treatment		Total	p value [H ₀ : M = W]
	Male Framing	Female Framing		
N	701 (50.2%)	695 (49.8%)	1,396 (100.0%)	
Gender				
Men	350 (49.9%)	345 (49.6%)	695 (49.8%)	0.914
Women	351 (50.1%)	350 (50.4%)	701 (50.2%)	
Age				
18	37 (5.3%)	29 (4.2%)	66 (4.7%)	0.525
19	47 (6.7%)	43 (6.2%)	90 (6.4%)	
20	55 (7.8%)	66 (9.5%)	121 (8.7%)	
21	91 (13.0%)	97 (14.0%)	188 (13.5%)	
22	126 (18.0%)	109 (15.7%)	235 (16.8%)	
23	122 (17.4%)	106 (15.3%)	228 (16.3%)	
24	116 (16.5%)	120 (17.3%)	236 (16.9%)	
25	107 (15.3%)	125 (18.0%)	232 (16.6%)	
Level of Education				
Less than a high-school diploma	6 (0.9%)	7 (1.0%)	13 (0.9%)	0.996
High-school diploma, no college	103 (14.7%)	100 (14.4%)	203 (14.5%)	
High-school diploma, currently in college	207 (29.5%)	207 (29.8%)	414 (29.7%)	
Associate degree	56 (8.0%)	56 (8.1%)	112 (8.0%)	
Bachelors degree	271 (38.7%)	274 (39.4%)	545 (39.0%)	
Graduate degree and higher	50 (7.1%)	43 (6.2%)	93 (6.7%)	
Other	8 (1.1%)	8 (1.2%)	16 (1.1%)	
Ethnicity simplified				
Asian	123 (17.5%)	128 (18.4%)	251 (18.0%)	0.571
Black	83 (11.8%)	93 (13.4%)	176 (12.6%)	
White	366 (52.2%)	351 (50.5%)	717 (51.4%)	
Mixed	73 (10.4%)	81 (11.7%)	154 (11.0%)	
Other	54 (7.7%)	40 (5.8%)	94 (6.7%)	
Prefer not to say	0 (0.0%)	1 (0.1%)	1 (0.1%)	

Notes: This table provides descriptives of the sample by treatment status (male vs. female framing). Percentages refer to the distribution of each variable within a column. The last column shows p-values of significant differences in means across treatment groups (M = male framing, W = female framing) computed using a Chi2-test for each variable.

Figure 4: Task Evaluation



Notes: This figure shows average task evaluations provided by study participants, by gender and task framing. Evaluations are elicited on three dimensions (task type, task performance and task enjoyment) shown in the different panels. Participants provide estimates of how they believe others perceive the task. All evaluations are measured via a 5-point Likert scale and range between the values -2 (very male-stereotypical evaluation) and +2 (very female-stereotypical evaluation); thus, positive values indicate female-stereotypical evaluations while negative values indicate male-stereotypical evaluations. P-values at the bottom of each graph were computed using Welch's two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average task evaluations between framings for the respective panel.

3.1.2 Task Evaluation

As described in section 2.2.1 to test if the male and female task framings indeed shift the perceived task stereotype, I analyse study participants' beliefs about other people's perceptions of the task (either male- or female-framed) as masculine/feminine, expecting men/women to perform better in the task and expecting men/women to enjoy working on the task more. Each panel of Figure 4 shows a different dimension on which participants had to evaluate the task. The two graphs in each panel show task evaluations separately for men and women.

The top most panel shows participants' evaluation of the task in terms of how masculine or feminine they believe other people perceive the task (in the following, this dimension is referred to as the "task type"). The evaluation is elicited on a five-point Likert scale ranging from very masculine (-2) to very feminine (+2). As shown in Figure 4, the framings work as intended: Participants on average believe that most other people consider the task more masculine in the male framing and more feminine in the female framing. The difference in the evaluation between male and female framing is highly significant with a p-value smaller than 0.0001 for both men and women.

Despite this, we can observe interesting gender differences in the levels of the task type perceptions. Participants expect others to perceive both framings as more stereotypical for their own gender (i.e., men believe others perceive both framings as more masculine, and women believe others perceive both framings as more feminine compared to the other gender). This finding is in line with related work. For instance, in Coffman (2014) participants rate all quiz categories as more gender-congruent for their own gender.⁴ As in my study, in Coffman (2014) the relative ratings are the same across gender and the differences occur only in levels. In addition, participants generally expect others to perceive the gender-congruent framing⁵ as more gender-stereotypical than opposite-gender subjects (for instance, men believe others perceive the task in the male framing as significantly more masculine compared to women and vice versa for the female framing). Lastly, participants report to believe that others perceive the gender-congruent task as more gender-stereotypical compared to the degree to which they expect others to perceive the gender-incongruent framing as counter-stereotypical (e.g., women believe the female-framed task is perceived by others as more

⁴Note that Coffman (2014) measures first-order beliefs.

⁵With gender-congruent I here refer to the gender of the study participant being in line with the framing and not to the gender of the other people whose task perceptions the participant is asked to report.

female-typed compared to the male-framed task being perceived as male-typed). All of these differences in task evaluations by gender occur not only for the first dimension of evaluating the task stereotype, but also the two other dimensions we consider (i.e., panels 2 and 3 in the Figure 4).

The second panel shows whether study participants think most other people believe men or women perform better in the respective task. Again, evaluations are elicited on a 5-point Likert scale ranging from -2 (“Men perform significantly better than women.”) to +2 (“Women perform significantly better than men.”). While men (left graph) evaluate the task as less extreme in this dimension compared to the task type dimension (especially in the female framing), the framings still shift task perceptions in the intended way. Note that in the female framing, men’s evaluations are not significantly different from zero, hence they believe others think men and women perform equally well in the task. Importantly, this is not a problem for identifying the effect of task stereotype in my experiment since the male framing is still perceived as significantly more “men-favoring” compared to the female framing. For women (right graph in panel 2), the difference in task perception between framings is highly significant as well and goes into the same direction as for men. However, it is driven by the contrary: In the male framing, women believe others think that men and women perform equally well, but in the female framing women report they expect most others to believe women perform significantly better in the task compared to men.

The last panel displays men’s and women’s evaluations of the task in terms of whether they believe others think men or women enjoy working on the task more. Again, the elicitation happens on a scale from -2 (“Men enjoy working on the task significantly more than women.”) to +2 (“Women enjoy working on the task significantly more than men.”). For both men (left graph) and women (right graph), the difference in evaluations goes in the expected direction and is highly significant for both genders.

Overall, the framings work on all three dimensions and significantly change how both male and female study participants expect others to evaluate the task.⁶

⁶Given that the task framings are a very subtle way to induce a change in task stereotype, the shift in perceived stereotype is significantly smaller compared to the existing literature. In [Coffman \(2014\)](#), for instance, the variation in task stereotype between the most male-typed and the most female-typed quiz categories is 2.74 standard deviations while in my study the variation between both framings is 0.461 SDs in the task type dimension, 0.268 SDs in the task performance dimension, and 0.309 SDs in the task enjoyment dimension. In section 3.4.2, I conduct a heterogeneity analysis splitting the sample based on whether participants perceive the framed task in the intended way to investigate treatment effects for a subsample of participants with a relatively strong induced variation in task stereotype by the framings.

3.1.3 Baseline Task Performance

Next, I investigate participants' task performance in the beginning of the experiment (in the following referred to as the baseline work round) depending on treatment (i.e., framing) assignment. Figure 5 depicts the distribution of the number of problems participants solved correctly when performing the matrix task at baseline, separately for men and women and by framing. On average, both men and women solved about 4 questions correctly, independent of the framing they were assigned to. The p-values depicted in black in the middle of both graphs refer to the difference in average performance within gender between framings and confirm that the performance differences between framings are insignificant for both sexes. This finding is noteworthy for at least two reasons: First, it contributes to a general discussion on the phenomenon of stereotype threat (e.g., [Steele et al. 2002](#)). Stereotype threat is usually defined as negative stereotypes about a group of people in a certain domain resulting in underperformance of these individuals in the respective field ([Steele et al. 2002](#)). While this idea has received significant attention in the literature in social psychology and behavioral economics over the recent decades, its robustness is still contested. My study does not provide evidence for the existence of stereotype threat. Second, the absence of performance differences by framing in my experiment mitigates concerns about performance differences confounding the main outcomes of interest in this study: beliefs about task ability (statistical self-stereotyping) and preference for the task (taste-based self-stereotyping).

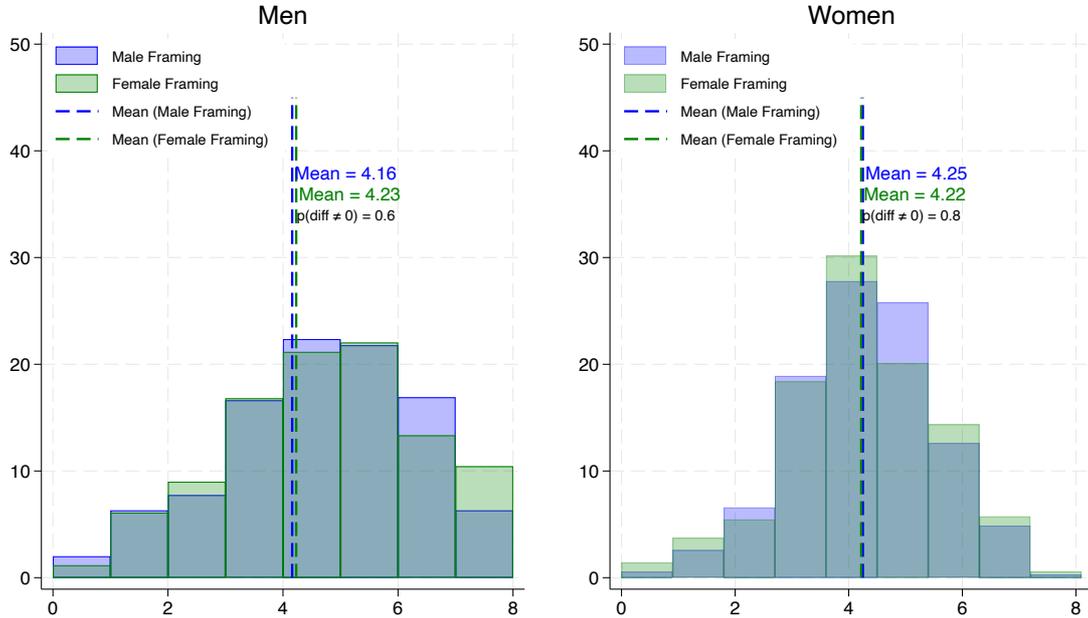
3.2 Main Analyses: Statistical Self-Stereotyping

In this project, statistical self-stereotyping is captured via the difference between treatment groups in stated beliefs about ability in the matrix task prior and post feedback provision. Beliefs are elicited on an 8-point scale indicating how many out of 8 matrix problems participants believe they have solved correctly in work round 1 (i.e., the baseline work round prior to feedback provision) or will solve correctly in work round 2 (post feedback).

3.2.1 Baseline Performance Beliefs (Prior to Performance Feedback)

We start with investigating baseline performance beliefs (Hypothesis 1.1). Figure 6 shows the distribution of the number of problems participants believe to have solved correctly when working

Figure 5: Baseline Task Performance for Men and Women

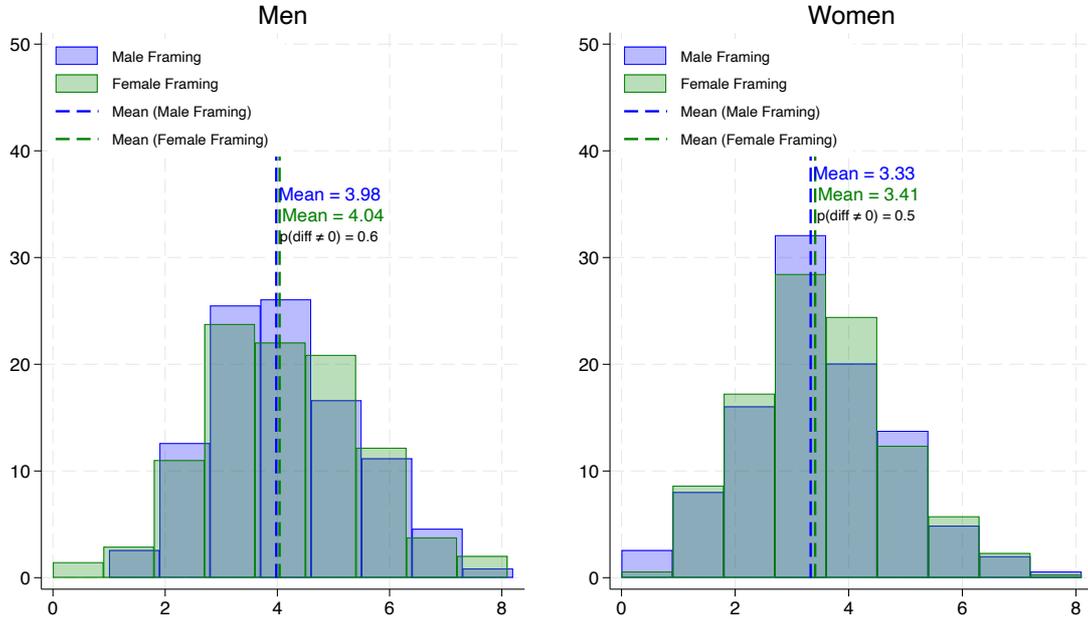


Notes: This figure depicts the distribution of baseline task performance in the matrix task by gender and task framing. The maximum achievable score is 8, the minimum achievable score is 0. P-values indicated in black in the middle of each graph were computed using Welch’s two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average baseline task performance across framings for the respective gender.

on the task in the first round, separately for men and women. As can be seen in the figure, men believe they have solved on average about 4 questions correctly both in the male and in the female framing, while for women the number is slightly lower with about 3.5 questions. Hence, women are overall less optimistic about their task performance than men. This is true independent of the task framing: the difference between framings is not statistically significant neither for men nor women as indicated by the p-values on the difference in average performance between framings depicted in black in the middle of both panels (0.6 for men and 0.5 for women). Note that men in my sample on average hold accurate beliefs about their past performance (they indeed solved on average 4 questions correctly at baseline as shown in Figure 5). Women, in contrast, are under-confident as they, on average, also solved 4 questions correctly in the baseline round but believe it was only 3.5.

Ex ante and particularly considering the previously mentioned concept of stereotype threat, it is unclear whether the assigned task stereotype also affects task performance. If this was the

Figure 6: Baseline Task Performance Beliefs for Men and Women



Notes: This figure depicts the distribution of baseline task performance beliefs in the matrix task by gender and task framing. The maximum achievable score is 8, the minimum achievable score is 0. P-values indicated in black in the middle of each graph were computed using Welch’s two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average baseline task performance beliefs across framings for the respective gender.

case, individuals might correctly perceive such differences in performance even prior to feedback, which could drive differences in beliefs about performance at baseline. As shown in section 3.1.3, on average subjects perform equally well in the matrix task independent of the framing assignment which mitigates such concerns. But given that average effects could conceal important heterogeneity across individuals, I am also exploring differences in beliefs about task performance prior to feedback conditioning on performance at baseline.

To analyse baseline task performance beliefs in more detail and, in particular, condition on baseline performance, I run simple regressions of baseline performance beliefs separately for men and women. Results are reported in Table 2. Columns (1) and (3) only condition on the treatment dummy **Male Framing** indicating whether the subject was assigned to the male-framed matrix task and demographic controls (age, ethnicity and education). In columns (2) and (4), I also condition on baseline task performance. For both genders, the framing does not significantly impact

Table 2: Baseline Performance Beliefs

	(1) Men	(2) Men	(3) Women	(4) Women
Male Framing	-0.0652 (0.117)	-0.0366 (0.108)	-0.0799 (0.111)	-0.0920 (0.0999)
Baseline Performance		0.362*** (0.0354)		0.457*** (0.0412)
Constant	2.814*** (0.594)	1.290*** (0.470)	3.725*** (0.640)	1.828** (0.738)
Controls	X	X	X	X
Observations	693	693	699	699
R-squared	0.038	0.181	0.041	0.229
Mean of Depend. Var.	4.013	4.013	3.366	3.366

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of baseline performance beliefs on the treatment indicator *Male Framing*. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, columns (2) and (4) control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 3 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

performance beliefs at baseline. The coefficient on the indicator for the male framing is negative for men and women for all specifications and slightly smaller for women which is directionally what we hypothesized. However, all coefficients on the indicator of the male framing are insignificant. On the other hand and as one would expect, baseline performance is significantly positively correlated with baseline performance beliefs. For women, baseline performance is more predictive of performance beliefs than for men. So overall, I conclude with the following result:

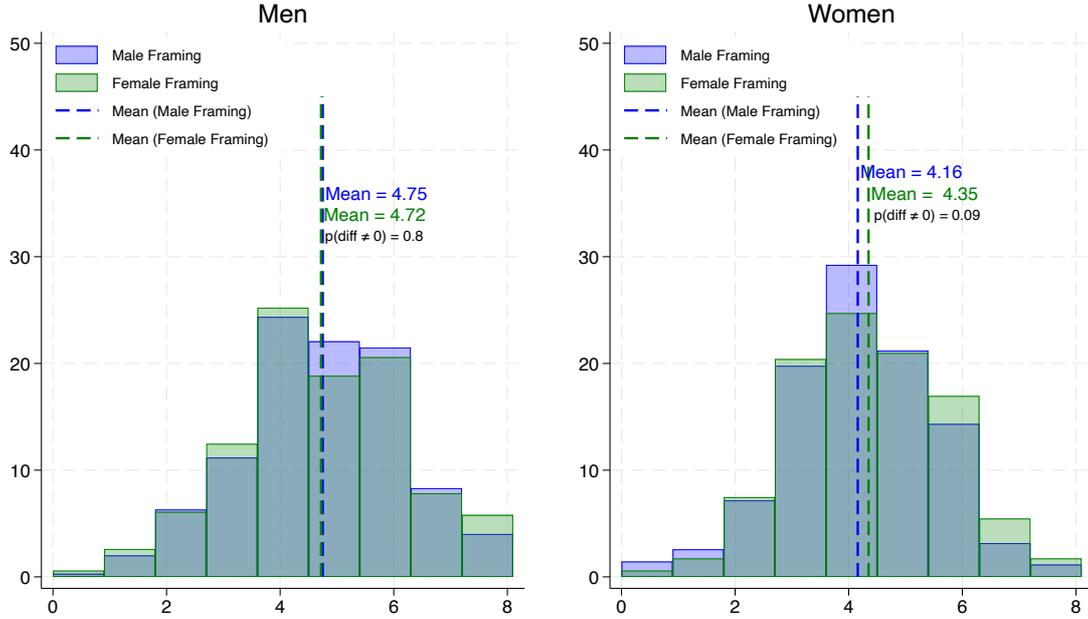
Result 1.1 *In terms of performance beliefs before feedback, I do not find evidence for statistical self-stereotyping in my experiment.*

3.2.2 Future Performance Beliefs (Post Performance Feedback)

Turning to Hypothesis 1.2, differential updating after feedback provision also provides evidence on statistical self-stereotyping: If updating differs by treatment, individuals place different weights on the provided feedback when forming their beliefs about future task performance depending on

the stereotype assigned to the task. Such differences in the difficulty to move beliefs across task stereotype indicate different degrees of strength of prior beliefs, hence providing additional evidence for the impact of stereotypes on beliefs about abilities (statistical self-stereotyping).

Figure 7: Future Task Performance Beliefs for Men and Women



Notes: This figure depicts the distribution of future task performance beliefs in the matrix task by gender and task framing. P-values in black in the middle of each graph were computed using Welch's two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average future task performance beliefs across framings for the respective gender.

As in the previous analysis, Figure 7 simply compares the distribution of future performance beliefs after feedback for men and women by treatment condition. For men, we again see no significant differences in their beliefs about their task performance in the future (i.e., the second work round) by treatment status. Overall, men hold more positive beliefs about their performance in the future compared to their beliefs about their past performance (Figure 6). Note that this does not stem from updating after feedback towards their true performance at baseline as beliefs about past performance are accurate. Hence, men believe they will perform better compared to their baseline performance if they do the task for a second round. For women, I do observe marginally significant differences in beliefs about future task performance by framing (p-value = 0.09). In particular, women are more optimistic about their performance in round 2 of

Table 3: Future Performance Beliefs

	(1) Men	(2) Men	(3) Women	(4) Women
Male Framing	0.0112 (0.123)	0.0598 (0.0972)	-0.213* (0.113)	-0.231*** (0.0851)
Baseline Performance		0.615*** (0.0324)		0.692*** (0.0399)
Constant	3.922*** (0.665)	1.331*** (0.442)	4.044*** (0.439)	1.173*** (0.400)
Controls	X	X	X	X
Observations	693	693	699	699
R-squared	0.044	0.409	0.046	0.468
Mean of Depend. Var.	4.733	4.733	4.258	4.258

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of future performance beliefs on the treatment indicator *Male Framing*. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, columns (2) and (4) control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 3 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

the task when they are assigned to the female compared to the male framing. Given that we do not see significant performance differences for women by framing, this is unlikely to be explained by differences in underlying true performance.

To investigate this in more detail, I run a regression analogous to the analyses in Table 2. Table 3 shows the results, again separately for men and women. As before, all model specifications include controls for age, ethnicity and education. In columns (2) and (4), I additionally control for baseline task performance. Note that participants have received feedback about their performance at baseline when forming their beliefs about task performance in the future. Hence, we observe that baseline performance strongly predicts future performance beliefs for both men and women. Also, we see that for men the framing does not significantly impact beliefs about future task performance. For women, however, being assigned to the male framing results in significantly less optimistic beliefs about their task performance in the future, conditional on true baseline performance (or, being assigned to the female framing significantly positively affects women’s beliefs about future task

performance). Even after learning their true performance in the task at baseline, framing the task as stereotypically masculine compared to stereotypically feminine has a significant and negative effect on future performance beliefs. Hence, I conclude:

***Result 1.2** For women, statistical self-stereotyping exists when it comes to the formation of beliefs about future task performance.*

3.3 Main Analyses: Taste-Based Self-Stereotyping

Next, I turn to the analyses of taste-based self-stereotyping. In my study, I attribute treatment effects in two main outcome measures to taste-based self-stereotyping: reported task enjoyment and a revealed preference measure for the task.

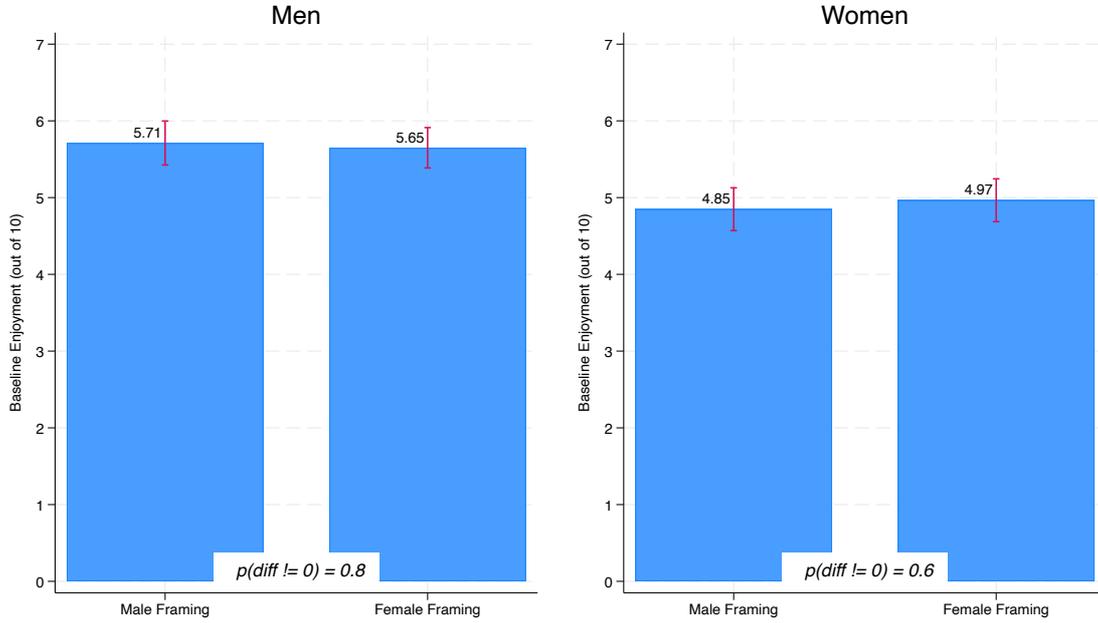
3.3.1 Task Enjoyment

Reported task enjoyment provides suggestive evidence on taste-based self-stereotyping: Does the task framing affect reported task enjoyment after baseline performance and expected task enjoyment for the second work round? Note that here it is crucial to control for beliefs about baseline performance to capture statistical self-stereotyping and identify the effect of task stereotypes on taste for the task that is additional to the impact on beliefs (especially, if we expect beliefs about performance and stated task enjoyment to be correlated). It is ex ante unclear whether controlling for true performance at baseline is necessary as well or if all impact of baseline performance on reported task enjoyment is already captured via the beliefs-channel.

Figures 8 and 9 provide first suggestive evidence on the Hypotheses 2.1 and 2.2 by showing reported task enjoyment for men and women by framing. While Figure 8 refers to enjoyment of the task in the baseline round which was elicited after baseline performance, Figure 9 depicts expected task enjoyment in the second round elicited prior to work round 2. Task enjoyment was measured on a 10-point Likert scale where 10 indicates the highest enjoyment level.

Overall, men on average report to enjoy working on the task more than women. This is true both for reported baseline enjoyment and expected future enjoyment. For men, we do not observe significant differences in reported enjoyment by task framing, neither in Figure 8 nor in Figure 9. Enjoyment varies within a small range between 5.65 and 5.75 for both baseline and future enjoyment

Figure 8: Baseline Task Enjoyment for Men and Women

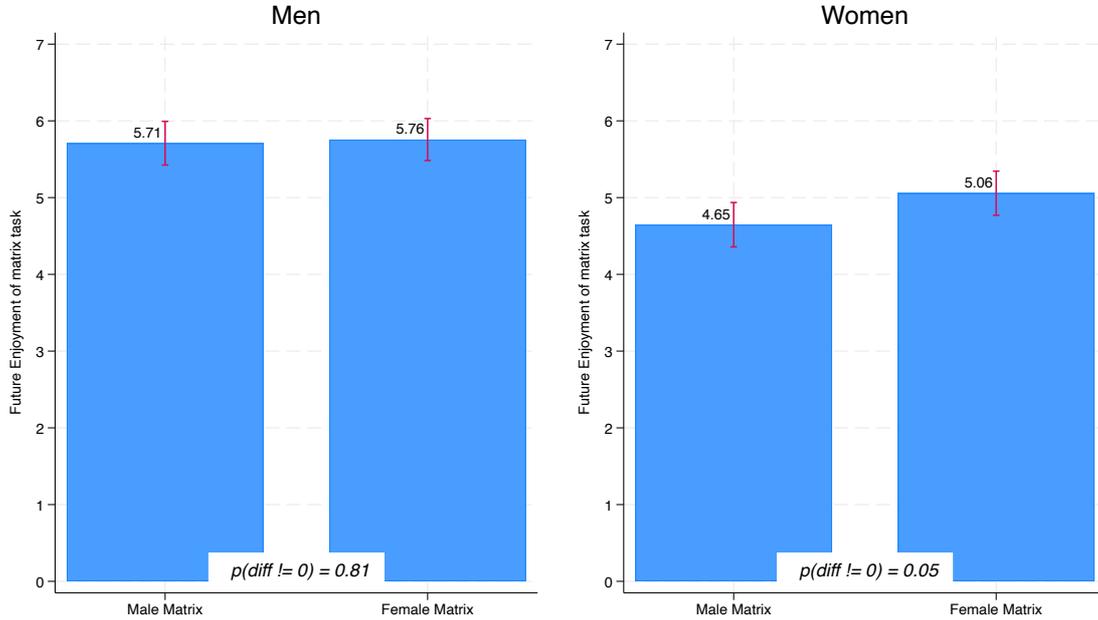


Notes: This figure depicts average reported baseline task enjoyment in the matrix task by gender and task framing. P-values at the bottom of each graph were computed using Welch’s two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average baseline task enjoyment across framings for the respective gender.

measures which, hence, are on average almost identical. Women on average report to enjoy the task less than men. They report a baseline enjoyment of around 4.85 in the female framing and 4.97 in the male framing. With a p-value of 0.6, this difference is insignificant. However, expected future enjoyment for women is significantly higher when they were assigned to the gender-congruent (i.e., female) framing compared to the gender-incongruent framing: Women expect an enjoyment of 5.06 in the second work round if they saw the female framing. However, if they were assigned to the male framing, expected enjoyment is only around 4.65. This difference is statistically significant with a p-value of 0.05. This provides suggestive evidence that, for women, the gender-stereotype assigned to the task not only matters for expected task performance in the future (see section 3.2) but also for expected task enjoyment.

To interpret this finding more rigorously, I control for task performance and performance beliefs to rule out that differences in enjoyment are driven by task ability. I do this in the form of a regression. Table 4 shows the results of a simple regression of reported baseline enjoyment on

Figure 9: Expected Future Task Enjoyment for Men and Women



Notes: This figure depicts average expected future task enjoyment in the matrix task by gender and task framing. P-values at the bottom of each graph were computed using Welch’s two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average expected future task enjoyment across framings for the respective gender.

the treatment indicator (“Male Framing”) and three different measures of (perceived) baseline performance.

Result 2.1 *The task stereotype has no significant impact on reported task enjoyment after the baseline round, conditional on beliefs about ability and past performance.*

Overall, we see that being assigned to the male framing relative to the female framing has no significant impact on reported task enjoyment after the baseline round in all specifications. Even though insignificant, the point estimates are consistent with the male framing having a stronger positive impact on reported baseline enjoyment for men compared to women. Columns (1) and (3) control for performance beliefs before feedback (i.e., how well participants think they have done in the task at baseline before they learn their true performance) and future performance beliefs after feedback. I find that both types of beliefs are very strongly positively correlated with baseline enjoyment. This makes sense: Participants’ subjective perception of how well they did in the task

Table 4: Task Enjoyment at Baseline

	(1)	(2)	(3)	(4)
	Men	Men	Women	Women
Male Framing	0.0655 (0.173)	0.0532 (0.173)	0.00770 (0.172)	0.00739 (0.174)
Past Performance Beliefs	0.501*** (0.0670)	0.504*** (0.0670)	0.621*** (0.0820)	0.621*** (0.0823)
Future Performance Beliefs	0.351*** (0.0691)	0.440*** (0.0804)	0.418*** (0.0792)	0.417*** (0.103)
Baseline Performance		-0.145** (0.0672)		0.00198 (0.0954)
Constant	1.511* (0.852)	1.767** (0.885)	-0.423 (1.303)	-0.426 (1.321)
Controls	X	X	X	X
Observations	693	693	699	699
R-squared	0.266	0.271	0.298	0.298
Mean of Depend. Var.	5.683	5.683	4.910	4.910

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of reported baseline enjoyment on the treatment indicator *Male Framing* and baseline and future performance beliefs. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, columns (2) and (4) control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 3 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

prior to receiving feedback as well as their beliefs on how well they will do in the future post feedback are two different metrics regarding task ability that subjects seem to take into account when reporting task enjoyment. In addition as indicated in column (2) baseline performance has a significant impact on reported task enjoyment for men. While the negative sign of the coefficient is surprising at first, it is important to point out that high correlations between performance beliefs and baseline performance make the interpretation of the point estimate difficult. Interestingly, for all three types of ability measures (baseline performance, baseline and future beliefs) the relationship of beliefs and reported enjoyment is stronger for women compared to men. Hence, it seems that for women (perceived) performance is more related to perceived enjoyment of the task.

Next, I explore expected future task enjoyment. Table 5 follows a similar structure as Table 4. The findings are qualitatively comparable to the findings in Table 4 and lead to the next result:

Table 5: Expected Future Task Enjoyment

	(1)	(2)	(3)	(4)	(5)	(6)
	Men	Men	Men	Women	Women	Women
Male Framing	-0.0511 (0.172)	-0.0671 (0.172)	-0.106 (0.115)	-0.244 (0.176)	-0.223 (0.176)	-0.229** (0.113)
Past Perf. Beliefs	0.197*** (0.0685)	0.202*** (0.0673)	-0.170*** (0.0501)	0.283*** (0.0829)	0.288*** (0.0832)	-0.202*** (0.0593)
Future Perf. Beliefs	0.682*** (0.0654)	0.797*** (0.0758)	0.473*** (0.0578)	0.802*** (0.0815)	0.881*** (0.103)	0.552*** (0.0732)
Baseline Performance		-0.189*** (0.0641)	-0.0818* (0.0436)		-0.128 (0.0909)	-0.130** (0.0574)
Baseline Enjoyment			0.738*** (0.0284)			0.789*** (0.0273)
Constant	1.198 (0.902)	1.531 (0.944)	0.226 (0.552)	-0.709 (1.525)	-0.514 (1.500)	-0.179 (0.641)
Controls	X	X	X	X	X	X
Observations	693	693	693	699	699	699
R-squared	0.297	0.306	0.692	0.318	0.320	0.727
Mean of Depend. Var.	5.729	5.729	5.729	4.854	4.854	4.854

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of expected future enjoyment on the treatment indicator *Male Framing* and baseline and future performance beliefs. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. In addition, columns (2) and (5) control for baseline performance and columns (3) and (6) control for reported baseline enjoyment. I exclude 3 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

Result 2.2 *The task stereotype has no significant impact on expected future task enjoyment, conditional on beliefs about ability and past performance.*

Different from the analyses of baseline task enjoyment, the role of future performance beliefs is significantly stronger for expected future enjoyment, whereas past performance beliefs play a smaller role. This is very reasonable as it suggests that (expected) performance in the particular work round considered is most relevant for (expected) enjoyment levels. Moreover, the regression results suggest that being assigned to the male framing instead of the female framing affects future enjoyment expectations more negatively for women compared to men for all model specifications, providing some supportive evidence for Hypothesis 2.2. However, the corresponding coefficient is

not significant in the pre-specified regressions controlling for beliefs about task ability and past performance only. One interesting alternative specification is shown in column (6) which controls for baseline enjoyment: Conditional on reporting the same enjoyment at baseline (and controlling for performance beliefs and baseline performance), women expect significantly lower enjoyment levels in the future work round when the task is counter-stereotypically framed. In a sense, this specification is very similar to the specification in columns (2) and (4) in Table 3: Controlling for the outcome of interest at baseline (i.e., baseline performance in Table 3 and baseline enjoyment in Table 5), the gender-incongruent framing significantly and negatively impacts women’s beliefs about future performance and women’s expectations regarding future enjoyment, respectively. It seems that once women are given some sort of leeway in the form of having to extrapolate from the baseline to a future work round, their beliefs both on abilities and enjoyment are significantly more pessimistic in the counter-stereotypical task framing. Hence, we find more evidence for Hypotheses 1.2 and 2.2 concerning the future work round compared to Hypotheses 1.1 and 2.1 on the baseline work round.

3.3.2 Revealed Preference Measure: Task Retake Decision

Given that self-reported liking of the task is impossible to incentivize and hence might be subject to various biases such as experimenter demand effects, another (behavioral) outcome to provide evidence on the importance of preferences in self-stereotyping is round 2 task choice. More precisely, the outcome of interest is the willingness to pay for retaking the matrix task instead of the outside option task. The underlying idea is the following: Measuring subjects’ beliefs about their future matrix task performance after obtaining perfect feedback on their baseline performance allows us to determine the retake choice that maximizes expected earnings E_i for each participant i in work round 2. Let p be the (constant) piece-rate payment for each correctly solved matrix problem in work round 2 and q_i individual i ’s belief of the number of matrix problems i will solve correctly in the second work round. Hence expected earnings of individual i in round 2 can be expressed as: $E_i = p \times q_i$. Furthermore, let $matrix_{ij}$ be a binary variable taking the value of 1 if individual i chooses the matrix task in row j of the MPL. An individual then maximizes expected earnings if (s)he applies the following decision rule:

$$matrix_{ij} = \begin{cases} 0, & \text{if } E_i \leq outsidepay_j \\ 1, & \text{if } E_i \geq outsidepay_j \end{cases}$$

where $outsidepay_j$ is a variable capturing the fixed payment offered for working on the outside option task in row j of the MPL. Following the above decision rule for the retake choice implies that an individual's choice can be perfectly explained by maximizing expected earnings. However, deviations from this behavior, can be interpreted as a revealed preference for (or against) the matrix task vis-à-vis the outside option, conditional on beliefs about ability (statistical self-stereotyping): If $matrix_{ij} = 1$ for $E_i \leq outsidepay_j$, then subject i expresses a revealed preference for the matrix task. If $matrix_{ij} = 0$ for $E_i \geq outsidepay_j$, then subject i expresses a revealed preference against the matrix task. Assuming at most one switching point for each study participant in the MPL, we can hence infer for every participant their revealed preference towards the matrix task conditional on statistical self-stereotyping working through beliefs about task ability. Differences in this measure by treatment constitute the treatment effect investigated in Hypothesis 2.3.

I will start the discussion of the results in this section by describing some patterns in participants' responses in the multiple price list (for a figure on the CDFs of switching behaviour of men and women on the MPL, please see appendix section B).⁷ First, switching happens all along the price list. This is not surprising as performance in the matrix task is also spread across all performance levels. Second, differences in switching behavior between genders are small. While women switch somewhat earlier towards the outside option task (i.e., at a somewhat lower outside pay), the differences are minor.

Next, we compare switching behavior across treatment conditions addressing the question whether subjects take different retake decisions depending on the framing they were assigned to. Figure 10a displays switching behavior along the MPL by treatment, separately for men and women. While some differences by treatment status emerge, these are small. Men switch slightly earlier towards the outside option task when assigned to the male framing throughout the price list, which can be

⁷Out of the 1'396 subjects in my study, only 127 show behaviours that are not monotonously increasing in the outside option pay (i.e., individuals that choose the outside option for $outsidepay_j = Y$ but not for $outsidepay_{i \neq j} > Y$). This show that the vast majority of participants displays rational behaviour when filling out the price list: If they choose the outside option task at $outsidepay_j = X$, they also choose the outside option for any $outsidepay_{i \neq j} > X$ and hence display at most one switching point along the MPL.

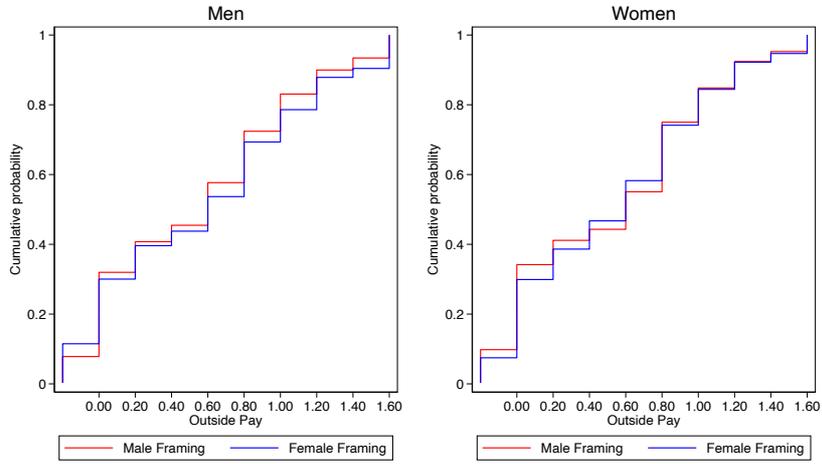
interpreted as showing a slight revealed preference for the matrix task when it is female framed. Women do this at low levels of outside pay, but show a slight preference for the matrix task in the male framing compared to the female framing at medium levels of pay for the outside option. At high levels of outside pay, no difference in revealed preference for the matrix task by framing occurs for women. Importantly, all these differences are small and the CDFs of switching behavior for the male and female framings do not appear/ significantly different from each other both for men and women.⁸ However, Figure 10a only shows the “raw” retake choices, not accounting for beliefs about future task performance. Given that we are interested in the effect of stereotypes on the revealed preference for the task conditional on performance beliefs, it is important to compare switching behavior along the MPL controlling for differences in beliefs about task ability by framings. I do this in Figure 10b.

Figure 10b again shows switching behavior on the MPL for men and women separately by assigned task framing. Importantly, the CDFs now display participants’ choices relative to the switching point on the price list that maximizes expected earnings. For this, I use beliefs about future task performance as measured before to calculate the point on the MPL where a subject should switch from the matrix task to the outside option to maximize expected earnings. For example, if a participant reported that she expects to solve 5 matrix problems correctly in work round 2, she should switch towards the outside option task as soon as the fixed payment for this task is larger than USD 1 (remember that each correctly solved matrix problem earns USD 0.20 and hence she expected to make at most USD 1 in the matrix task). In Figure 10b, a switching point of 0 indicates switching exactly at the point on the MPL that maximizes expected earnings in work round 2, switching at a point < 0 implies switching prior to this point (hence, displaying a revealed preference against the matrix task) and switching at a point > 0 means switching later than this point (hence, displaying a revealed preferences for the matrix task).

When comparing Figures 10a and 10b for men, the interpretation stays qualitatively the same:

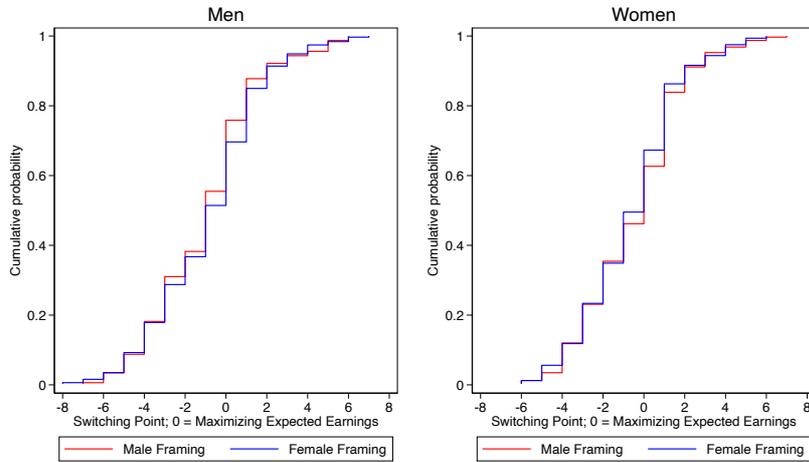
⁸It is interesting to see whether switching behavior correlates with reported task enjoyment at baseline and expected task enjoyment in the future in ways that differ by gender-congruency of the task framing. Generally, correlations (reported in Appendix section B) are small and insignificant, especially for men and women seeing the female framing. For the male framing, we observe a slightly significant positive correlation for women and a significant negative correlation for men when considering the correlation of expected future enjoyment and retake behavior: The more women (men) report to expect to enjoy the task in work round 2, the more (less) they choose it for retake. This goes against our initial hypotheses of gender-congruency making the task more enjoyable and leading to higher retake choices.

Figure 10: CDF of Switching along MPL
(a) By Task Framing



Notes: The two graphs in Panel (a) of this figure show the cumulative distribution functions of the switching points along the multiple price list, separately for men and women and by assigned task framing. The cumulative probability is depicted on the y-axis, the fixed payment for the outside option as it increases along the MPL is shown on the x-axis. The figure excludes 127 observations with non-unique switching points along the MPL.

(b) Relative to Expected Future Performance



Notes: The two graphs in Panel (b) of this figure show the cumulative distribution functions of the switching points along the multiple price list relative to the switching point that would maximize expected earnings in work round 2 according to participants stated beliefs about their future performance, separately for men and women and by assigned task framing. The cumulative probability is depicted on the y-axis, the switching point is shown on the x-axis. Here, the value of 0 describes the point at which participants should switch if they were maximizing expected earnings; switching points > 0 describe switching points where participants express a preferences for the matrix task; switching points < 0 describe switching points where participants express a distaste for the matrix task. The figure excludes 127 observations with non-unique switching points along the MPL.

While differences are small in magnitude, it seems that men switch slightly earlier to the outside option when seeing the gender-congruent framing. The similarity of the two figures for men makes sense, given that framings did not affect beliefs about future task performance for male participants. For women, the CDFs change somewhat comparing Figures 10a and 10b: Analogous to men, it now appears that women prefer the matrix task to a larger degree when seeing the gender-incongruent framing. However, differences in CDFs between framings are still small and not statistically significant. Hence, I conclude with the last result:

Result 2.3 *There is no strong evidence on taste-based self-stereotyping neither for men nor for women in my study.*

If anything, it appears that both genders display a preference for the counter-stereotypical task at retake which could be an interesting pattern to be explored in future research.

3.4 Heterogeneity Results

This section presents several heterogeneity results. These were not pre-registered and hence need to be interpreted with caution. Still, they provide interesting insights into the previously discussed findings.

3.4.1 Statistical Self-Stereotyping, Baseline Performance and Belief Updating

As shown in section 3.2, beliefs about future task performance are significantly and negatively affected by the counter-stereotypical task stereotype for women. In this section, I am pinning down more precisely which women drive this result. First, I consider heterogeneity by baseline task performance as it seems plausible that the impact of gender stereotypes on beliefs about task ability varies with prior task performance. I define low baseline performance as performing average or worse in the baseline work round (i.e., solving at most 4 matrix problems correctly). Performing above average at baseline (i.e., solving more than 4 matrix problems correctly) is defined as high baseline performance. I find that the effect of task stereotype on beliefs about future task performance is entirely driven by high-performing women.⁹ For details on the estimation and results, please refer to appendix section C.1.

⁹As suggested by the main results on statistical self-stereotyping, the task stereotype does not affect beliefs about future task performance for men, independent of their baseline performance.

We know from existing literature that gender-stereotypes matter for belief-updating (Coffman et al., 2021b, 2024). This suggests belief updating as a channel through which statistical self-stereotyping occurs. To investigate this hypothesis, I divide my sample based on the performance feedback participants receive after the baseline work round. A participant receives negative feedback if her reported belief about her baseline performance is higher than her true baseline performance, such that the performance feedback reveals a negative piece of information to the participant. In contrast, a participant receives positive performance feedback if his true performance at baseline is higher than his previously stated belief. I keep the sample split by baseline performance (low vs. high) to analyze interaction effects of baseline performance and performance feedback in statistical self-stereotyping. My findings show that the task stereotype matters for women (and men!) that perform above average at baseline and receive positive performance feedback: For those individuals, the counter-stereotypical task stereotype has a highly significant, negative impact on future performance beliefs. For women, the effect is larger than the previously detected effects for high performing women and women in general. This result is particularly noteworthy as it occurs for both women and men, suggesting that also the remaining effects I have so far detected for women could also exist for men but might be too small to be detected with my current sample. Again, details on the estimation results can be found in appendix C.1.¹⁰

3.4.2 Sample Split: Successful Task Framings

As suggested in the previous section, it is possible that some effects I investigate in this study are too small to be detected with my current sample size (especially for men). For this reason, I now turn to another set of analyses that appears highly valuable in gaining insights into statistical and taste-based self-stereotyping, especially considering the potential existence of small effects, non-detectable with the previous analyses. More precisely, I conduct a median-split of the main sample in terms of task evaluation (see section 3.1.2). Given the subtle nature of my experimental manipulation (male vs. female task framing), it is possible that effects are non-detectable due to the relatively small impact task framings have on the perceived task stereotype. Hence, in the following I present the

¹⁰Note that the likelihood to receive positive vs. negative feedback varies with baseline performance: For instance, subjects that perform high at baseline mechanically have more room to be too pessimistic about their performance and are hence more likely to receive positive feedback. However, the fact that the average participant (man and woman) in my sample answers around 4 questions correctly and only very few subjects score extremely low or high in the matrix task in both framings mitigates concerns that my findings are driven by ceiling or floor effects.

results of analyses that split the sample according to the strength of participants' task evaluations.

To split the sample as described above, I am generating a summary variable that averages each participant's task evaluation across the three dimensions I am using in this study: task type, task performance and task enjoyment. I then assign all participants with a summary score above the median (calculated within each framing) in terms of perceiving the framed task in the intended way (i.e., expecting others to evaluate the male-framed task as masculine/men performing better than women/men enjoying the task more than women and vice versa for the female-typed task) to the sample "Stereotypical" and those below or at the median to the sample "Non-Stereotypical". I expect statistical and taste-based self-stereotyping to only occur for participants in the "Stereotypical" subsample given that only for them the framings work as intended. For details on assigning participants to subsamples, please refer to appendix section [C.2](#).

I find that in the "Stereotypical" subsample, not only women's beliefs about their future performance (post feedback) but also women's baseline performance beliefs (prior to feedback) are significantly and negatively impacted by the male task framing. Further, men's expected future performance is also significantly impacted by the male framing in this subsample but in a positive way, hence providing evidence for statistical self-stereotyping for men. Lastly, I find that for female participants in the "Stereotypical" subsample, the male framing has a significant and negative impact on both baseline and expected future task enjoyment, providing evidence for hypotheses 2.1 and 2.2 for women. As expected, I do not detect any effects of task framings and hence stereotypes on outcomes considered in the "Non-Stereotypical" subsample. More details on the subsample analyses can be found in appendix section [C.2](#).

4 Conclusion

To my knowledge, this study is the first to differentiate and measure two types of self-stereotypical behavior: statistical and taste-based self-stereotyping. It does so using a novel experimental design which relies on gendered task framings to exogenously induce a shift in the perceived gender stereotype of a task while holding the task itself fixed. This naturally extends designs that have been used in the existing literature and that rely on different tasks or task domains (one male-typed and one female-typed) to shift the task stereotype in the experiment. Keeping the task substantively

the same, the design of my experiment improves the identification of self-stereotyping, especially by mitigating concerns regarding omitted variables like task characteristics that change with the task (domain) and that are correlated both with the task stereotype and the outcome of interest.

I find that being assigned to the task in the male framing compared to the female framing significantly lowers beliefs about future task performance for female participants, even conditional on task performance and feedback. This provides evidence for statistical self-stereotyping for women. I do not find strong support for the existence of taste-based self-stereotyping in the main sample of my study, even though the male task stereotype negatively affects women's expected future task enjoyment. Heterogeneity analyses show some interesting, more nuanced results.

There are many interesting avenues for future research building on this paper. For instance, it would be interesting to change the setting and see if statistical and especially taste-based self-stereotyping happen in slightly different but related contexts. For example, making the retake choice observable to peers could be an important factor leading to taste-based self-stereotyping as preferences for stereotype compliance are likely stronger in social settings. Another possible extension would be to consider a group outcome (e.g., group performance in the task). Important papers on self-stereotyping such as [Coffman \(2014\)](#) and [Coffman et al. \(2021a\)](#) investigate self-stereotypical behavior in a group context. I abstract away from this in my study to measure statistical and taste-based self-stereotyping in a very simple setting first where individuals are not interacting with others and can only signal to themselves by complying with stereotypes. But it is possible (and likely) that comparing yourself to others - as it happens in group contexts - is an important factor determining self-stereotypical behavior.

Overall, this study rather serves as a benchmark for statistical and taste-based self-stereotyping by focusing on a context where the impact of stereotypes is limited through the rather narrow design. I consider this as a starting point for investigating self-stereotyping, and especially the two types of statistical and taste-based self-stereotyping, more closely in future work.

References

- Alan, S., Ertac, S., Kubilay, E., and Loranth, G. (2020). Understanding gender differences in leadership. *The Economic Journal*, 130(626):263–289.
- Barron, K., Schweighofer-Kodritsch, S., and Stelzeneder, D. (2023). Gender differences in competitiveness: Does the perceived gender domain of the task matter? experimental evidence. *unpublished manuscript*.
- Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.
- Brenøe, A. A. (2022). Brothers increase women’s gender conformity. *Journal of Population Economics*, 35(4):1859–1896.
- Brenøe, A. A. and Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4):1009–1054.
- Bursztyn, L., Fujiwara, T., and Pallais, A. (2017). ‘acting wife’: Marriage market incentives and labor market investments. *American Economic Review*, 107(11):3288–3319.
- Chen, J. and Houser, D. (2019). When are women willing to lead? the effect of team gender composition and gendered tasks. *The Leadership Quarterly*, 30(6):101340.
- Coffman, K., Collis, M. R., and Kulkarni, L. (2024). Stereotypes and belief updating. *Journal of the European Economic Association*, 22(3):1011–1054.
- Coffman, K., Flikkema, C. B., and Shurchkov, O. (2021a). Gender stereotypes in deliberation and team decisions. *Games and Economic Behavior*, 129:329–349.
- Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.
- Coffman, K. B., Araya, P. U., and Zafar, B. (2021b). A (dynamic) investigation of stereotypes, belief-updating, and behavior. Technical report, National Bureau of Economic Research.
- Dreber, A., Von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, 17:154–172.

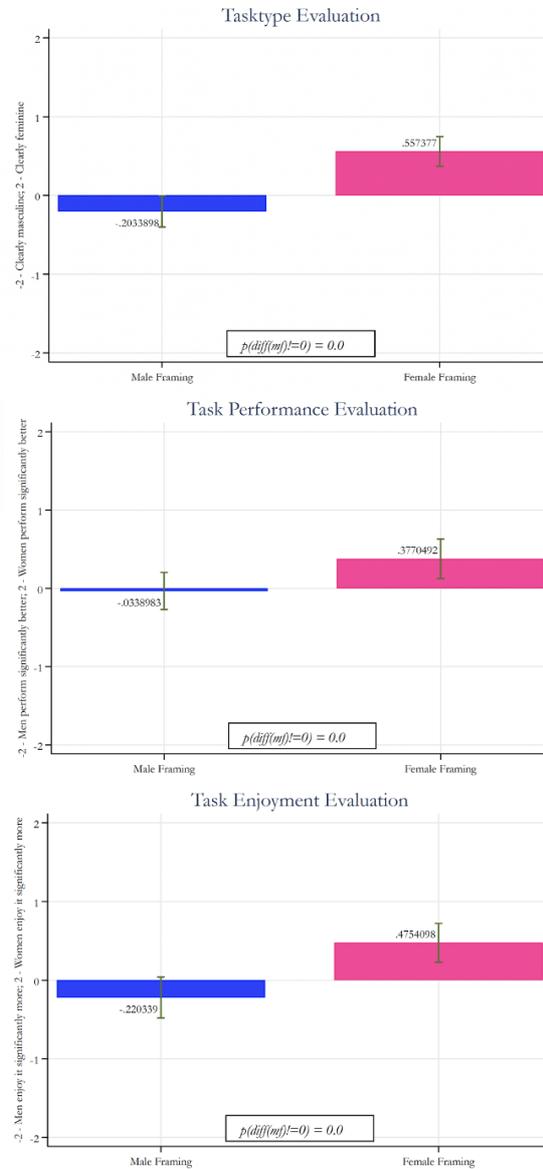
- Exley, C. L. and Kessler, J. B. (2022). The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381.
- Grosse, N., Riener, G., and Dertwinkel-Kalt, M. (2014). Explaining gender differences in competitiveness: Testing a theory on gender-task stereotypes. *Available at SSRN 2551206*.
- Kamas, L. and Preston, A. (2012). Are women really less willing to compete than men? gender stereotypes, confidence, and social preferences. Technical report, Mimeo Haverford College.
- Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.
- Rosenberg, B. D. and Siegel, J. T. (2018). A 50-year review of psychological reactance theory: Do not read this article. *Motivation Science*, 4(4):281.
- Steele, C. M., Spencer, S. J., and Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In *Advances in experimental social psychology*, volume 34, pages 379–440. Elsevier.
- Zölitz, U. and Feld, J. (2021). The effect of peer gender on major choice in business school. *Management Science*, 67(11):6963–6979.

Appendix

A Pre-Testing

Figure 11 shows results from pretesting the task framing. The pre-test was done with 120 individuals, out of which 59 were assigned to the male framing, and 61 to the female framing. Task perception was elicited exactly as in the main study; hence, please refer to section 2.2.1 for a detailed explanation of the elicitation of task perceptions. In figure 11, each panel depicts one task evaluation dimension (task type, i.e., whether individuals believe others perceive the task as rather masculine or feminine; task performance, i.e., whether individuals believe others expect men or women to perform better in the task; and task enjoyment, i.e., whether individuals believe others expect men or women to enjoy working on the task more). The significant differences in task perception between framings in all three dimensions show that the framings successfully induce a change in the perceived task stereotype in the pre-test.

Figure 11: Task Evaluation by Framing (Pretesting)



Notes: This figure shows average task evaluations provided by pretest participants by task framing. Evaluations are elicited on three dimensions (task type, task performance and task enjoyment) shown in the different panels. Participants provide estimates of how they believe others perceive the task. All evaluations are measured via a 5-point Likert scale and range between the values -2 (very male-stereotypical evaluation) and +2 (very female-stereotypical evaluation); thus, positive values indicate female-stereotypical evaluations while negative values indicate male-stereotypical evaluations. P-values at the bottom of each graph were computed using Welch's two-sample t-test, which does not assume equal variances between groups. They show the significance of the difference in average task evaluations between framings for the respective panel.

B Additional Tables and Figures

B.1 Correlations of Reported Enjoyment and Retake Behavior

Table 6: Correlations of Reported Enjoyment and Retake Behavior

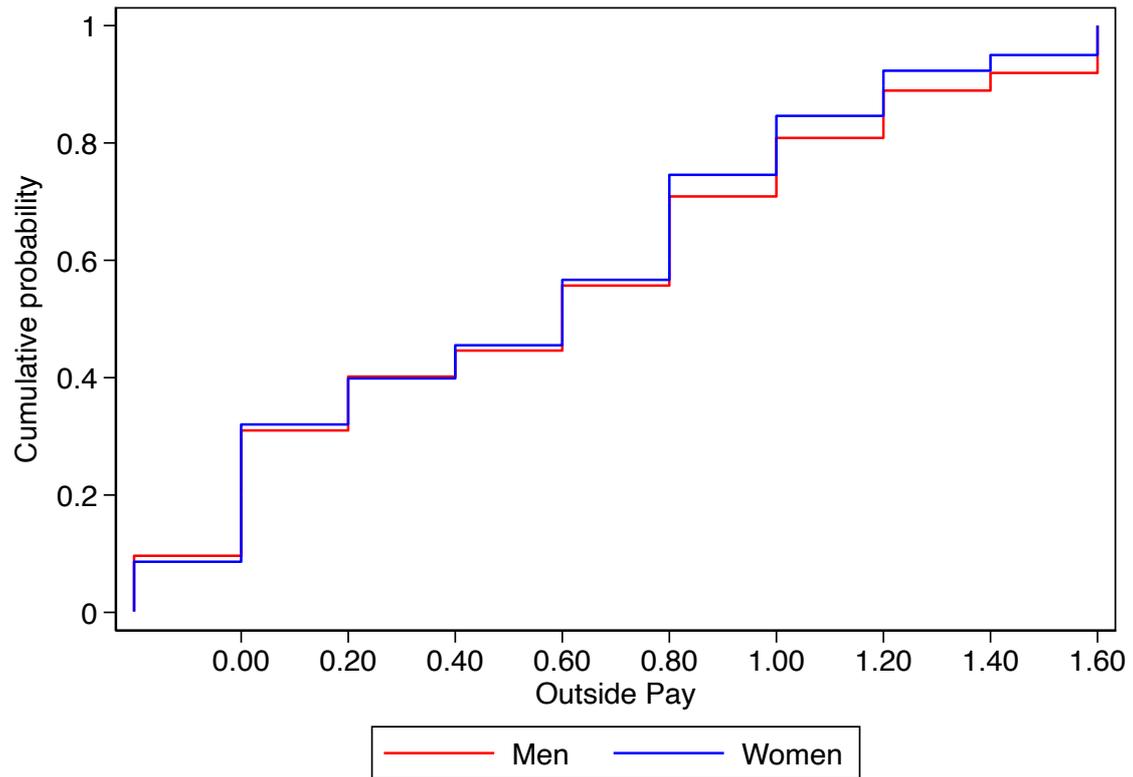
	Correlations with MPL Switching Point	
	Men (1)	Women (2)
Baseline Enjoyment Male Framing	-0.081	0.043
Baseline Enjoyment Female Framing	-0.047	-0.049
Expected Future Enjoyment Male Framing	-0.141**	0.098*
Expected Future Enjoyment Female Framing	-0.017	-0.050

Notes: Table shows Pearson correlations of switching behavior on the multiple price and (i) reported task enjoyment at baseline (rows 1, 2) (ii) expected task enjoyment for work round 2 (rows 3, 4). Significance of correlations (based on t-tests) indicated by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.2 Cumulative Distribution Function of Switching along MPL

Figure 12 shows the cumulative distribution functions of switching points along the multiple price list by gender.

Figure 12: CDF of Switching along the MPL for Men and Women



Notes: This figure shows the cumulative distribution functions of the switching points along the multiple price list, separately for men and women. The cumulative probability is depicted on the y-axis, the fixed payment for the outside option as it increases along the MPL is shown on the x-axis. The figure excludes 127 observations with non-unique switching points along the MPL.

C Heterogeneity Analyses

C.1 Statistical Self-Stereotyping, Baseline Performance and Belief Updating

The following analyses consider future performance beliefs for men and women by baseline task performance. Table 7 presents the results of regressions of future task performance beliefs on the treatment indicator *Male Framing* (columns (1) and (3)), controlling for baseline task performance (columns (2) and (4)), for individuals that perform average or worse at baseline (i.e., solving at most 4 matrix problems correctly). The regressions show that for low performing women seeing the male framing compared to the female framing has no significant impact on their beliefs about their future task performance.

Hence, the treatment effect of seeing the male-typed task on future performance beliefs for women must be driven by those women that perform very well at baseline (i.e., solving more than 4 matrix problems correctly and hence performing better than the average). We confirm this in table 8 which shows regressions analogous to those in table 7. Even conditional on baseline performance, high-performing female subjects who work on the task with the male stereotype attached to it hold significantly lower beliefs about their future performance compared to women performing equally well at baseline but working on the female-framed task. Given that this subsample drives the main effect observed for women in the full sample, the point estimate of the coefficient of *Male Framing* is significantly larger for this sample: When controlling for baseline performance, the difference in expected future performance by task framing for high-performing women is around 1/3 of a question and hence quantitatively meaningful. For men, the framings do not matter for expected future task performance neither for low nor for high performing men at baseline as seen in the first two columns of tables 7 and 8.

Next, to investigate whether the task stereotype matters for belief-updating, I analyze heterogeneity in the impact of task stereotype on future performance beliefs depending on the type of feedback a participant receives after performing the task at baseline. A participant receives negative feedback if her belief about her baseline performance is higher than her actual performance. In this case, the feedback she gets about her past performance is negative relative to her stated belief before feedback. In contrast, a participant receives positive feedback if her reported belief about baseline performance is lower than her true performance.

Table 7: Future Performance Beliefs by Framing (Low Performing Individuals)

	(1) Men	(2) Men	(3) Women	(4) Women
Male Framing	-0.0119 (0.151)	0.00628 (0.143)	-0.118 (0.124)	-0.124 (0.112)
Baseline Performance		0.413*** (0.0748)		0.520*** (0.0950)
Constant	2.778*** (0.405)	1.405*** (0.421)	3.901*** (0.550)	2.149*** (0.480)
Constant	X	X	X	X
Observations	379	379	405	405
R-squared	0.109	0.197	0.030	0.190
Mean of Depend. Var.	3.931	3.931	3.536	3.536

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of future performance beliefs on the treatment indicator *Male Framing* for individuals that have solved at most 4 matrix problems correctly in the baseline work round. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, columns (2) and (4) control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 2 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

Tables 9 and 10 show the results of regressions of future performance beliefs on the male task stereotype (*Male Framing*), controlling for baseline performance. The regressions are run separately by gender and, in addition, by baseline performance (high vs. low as defined above). The results clearly show that the significant negative impact of the gender-incongruent task stereotype on expected future performance for women is driven by high performing women receiving positive feedback about their baseline performance. With an increase in expected future performance of about 2/5 of a question, the effect is even larger than the one for high performing women in general. Interestingly, the effect is mirrored by high performing male respondents receiving positive feedback: For them, seeing the gender-congruent (i.e., male) framing compared to the female one, significantly increases their beliefs about task performance in the future. This finding suggests that for men, statistical self-stereotyping might exist similarly as for women but with more noise making is harder to detect statistically significant effects.

Table 8: Future Performance Beliefs by Framing (High Performing Individuals)

	(1) Men	(2) Men	(3) Women	(4) Women
Male Framing	0.0486 (0.147)	0.138 (0.131)	-0.482*** (0.149)	-0.356*** (0.128)
Baseline Performance		0.751*** (0.0805)		0.831*** (0.0817)
Constant	6.356*** (0.497)	1.801* (0.975)	4.235*** (0.497)	0.0618 (0.623)
Controls	X	X	X	X
Observations	314	314	294	294
R-squared	0.036	0.259	0.094	0.334
Mean of Depend. Var.	5.701	5.701	5.252	5.252

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of future performance beliefs on the treatment indicator *Male Framing* for individuals that have solved more than 4 matrix problems correctly in the baseline work round. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, columns (2) and (4) control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 1 observation that has a non-valid value for ethnicity.

C.2 Sample Split: Successful Task Framings

Given that the task framings induce the shift in task stereotype very subtly, the difference in perceived task stereotype across framings is rather small compared to the existing literature using different tasks or task domains to vary the gender stereotype within the experiment. To further investigate my hypotheses on statistical and taste-based self-stereotyping, I will now split the sample based on participants' perceived task stereotype (discussed in section 3.1.2). I rerun all analyses for the subsample of participants that evaluate the task with the respective framing in the intended way, i.e., perceiving the task with the male (female) framing as stereotypically male (female) (and for the subsample of participants that do not). This allows to investigate my hypotheses on statistical and taste-based self-stereotyping conditional on the experimental design (i.e., using task framings to exogenously change task stereotypes) working particularly well and leverage a larger shift in task stereotype induced by the framing.

To split the sample as described above, I am generating a summary variable that averages

Table 9: Future Performance Beliefs by Framing (Negative Feedback Participants)

	(1)	(2)	(3)	(4)
	Men		Women	
	Low Perf.	High Perf.	Low Perf.	High Perf.
Male Framing	0.0313 (0.169)	-0.0111 (0.223)	-0.108 (0.165)	-0.0660 (0.203)
Baseline Performance	0.542*** (0.0815)	1.018*** (0.165)	0.531*** (0.122)	0.818*** (0.114)
Constant	1.463*** (0.561)	0.990 (1.109)	2.301*** (0.518)	1.230** (0.605)
Constant	X	X	X	X
Observations	275	107	205	58
R-squared	0.277	0.418	0.243	0.690
Mean of Depend. Var.	4.149	6.121	3.698	5.983

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of future performance beliefs on the treatment indicator *Male Framing* for individuals receiving negative performance feedback after the baseline work round, i.e., their stated beliefs about their baseline performance are higher than their actual baseline performance. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, I control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 1 observation that has a non-valid value for ethnicity.

each participant’s task evaluation across the three dimensions I am using in this study.¹¹ I then assign all participants with a summary score above the median (calculated within each framing) in terms of perceiving the framed task in the intended way to the sample “Stereotypical” and those below or at the median to the sample “Non-Stereotypical”. Table 11 shows summary statistics of the summary score by framing. As can be seen from the table, the framings on average work as intended. The male-framed task on average receives a negative evaluation score (-0.093) indicating a male-stereotypical task perception, while the female-framed task on average receives a positive evaluation score (0.225) indicating a female-stereotypical task perception. The female framing is more successful on average in achieving the intended task perception, considering the larger absolute magnitude in the average task evaluation. Correspondingly the median summary task

¹¹The three dimensions are 1) task type: how masculine/feminine do participants expect others to perceive the task, 2) task performance: do participants expect others to believe that men or women perform better in the task, 3) task enjoyment: do participants expect others to believe that men or women enjoy working on the task more? For further details, see section 3.1.2.

Table 10: Future Performance Beliefs by Framing (Positive Feedback Participants)

	(1)	(2)	(3)	(4)
	Men		Women	
	Low Perf.	High Perf.	Low Perf.	High Perf.
Male Framing	-0.0154 (0.231)	0.415** (0.164)	-0.131 (0.147)	-0.385*** (0.142)
Baseline Performance	0.729*** (0.157)	0.812*** (0.0855)	1.036*** (0.101)	0.924*** (0.0908)
Constant	0.332 (0.553)	1.192 (0.979)	-0.0321 (0.658)	-0.721 (0.608)
Controls	X	X	X	X
Observations	104	207	200	236
R-squared	0.281	0.330	0.401	0.382
Mean of Depend. Var.	3.356	5.483	3.370	5.072

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of a regression of future performance beliefs on the treatment indicator *Male Framing* for individuals receiving positive performance feedback after the baseline work round, i.e., their stated beliefs about their baseline performance are lower than their actual baseline performance. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). In addition, I control for baseline performance. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 2 observations that have non-valid values for ethnicity and 1 observation that prefers not to state their ethnicity in the survey.

evaluation score for the female framing (0.33) is larger than the median value for the male framing (0). For splitting the sample at the median value, I decided to include the median value in the “Non-Stereotypical” subsample. This ensures that the “Stereotypical” subsample contains those observations that are most “convinced” by the task-framings.

Table 11: Summary Statistics Task Evaluation Summary Score

	Male Framing	Female Framing
Mean	-0.093	0.225
Std. dev.	0.762	0.733
Median	0	0.333

Notes: This table shows summary statistics of the summary evaluation score which aggregates the task evaluations for each participant across the three evaluation dimensions by framing. Values below zero indicate a valuation that is on average more male-stereotyped, values above zero indicate a valuation that is on average more female-stereotyped.

C.2.1 Statistical Self-Stereotyping

Tables 12 and 13 show the results of regressions of baseline performance beliefs and future performance beliefs on task stereotype (*Male Framing*) and baseline performance by gender for the previously defined subsamples “Non-Stereotypical” and “Stereotypical”.

The findings support the claim that stereotypes are indeed the driving force behind the significant negative impact of the male framing on future performance beliefs for women. This becomes clear by comparing column (4) of both tables: Only for women who perceive the task in the intended way the effect of *Male Framing* on future performance beliefs appears. Further evidence for the stereotype-channel is provided when comparing column (3) in table 12 with column (3) in table 13: In the subsample of men who perceive the task in the intended, stereotypical way, we find a quantitatively meaningful, highly significant, and positive effect of the male framing on future performance beliefs. Hence, for both women and men seeing the gender-congruent framing significantly and positively impacts beliefs about future task performance, controlling for past performance. This finding further supports the hypotheses stated in section C.1 that statistical self-stereotyping also occurs for men but is harder to detect.

Lastly, table 13 also shows a significant and negative coefficient for *Male Framing* on baseline performance beliefs for women. Hence, at least for women and conditional on perceiving the task in the intended way, the gender-incongruent task stereotype even impacts beliefs about past performance prior to feedback negatively.

C.2.2 Taste-Based Self-Stereotyping

As a first indicator of taste-based self-stereotyping, I again start by investigating the impact of task stereotype on reported baseline and expected future task enjoyment. The results of the corresponding regressions for the “Non-Stereotypical” and “Stereotypical” subsamples are reported in tables 14 and 15.

As for statistical self-stereotyping, the tables show no impact of the male task framing on enjoyment measures for the subsample of participants who do not evaluate the framed tasks in the stereotypical, intended way. However, for the subsample of individuals that do, we see significant and negative effects of the male framing on reported task enjoyment for women in all specifications

Table 12: Performance Beliefs by Framing (Non-Stereotypical Subsample)

	(1)	(2)	(3)	(4)
	Baseline Performance Beliefs		Future Performance Beliefs	
	Men	Women	Men	Women
Male Framing	0.0242 (0.141)	0.0496 (0.134)	-0.0807 (0.129)	0.0215 (0.113)
Baseline Performance	0.349*** (0.0473)	0.395*** (0.0607)	0.581*** (0.0434)	0.668*** (0.0579)
Constant	2.285*** (0.422)	1.685* (0.865)	2.346*** (0.618)	1.004** (0.485)
Controls	X	X	X	X
Observations	414	393	414	393
R-squared	0.183	0.213	0.397	0.462
Mean of Depend. Var.	4.053	3.326	4.734	4.242

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of regressions of baseline performance beliefs (columns (1) and (2)) and future performance beliefs (columns (3) and (4)) on the treatment indicator *Male Framing* and baseline performance by gender and for individuals in the “Non-Stereotypical” subsample, i.e., participants whose stated task evaluations do not correspond to the intended task stereotype. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 1 observation that prefers not to state their ethnicity in the survey.

of table 15. These results again provide strong evidence that it is the task stereotype that drives the impact of the male framing on women’s expected future task enjoyment discussed in section 3.

Lastly, I analyze the revealed preference measure for the matrix task for the “Non-Stereotypical” and “Stereotypical” subsamples.

Figures 13 and 14 depict the cumulative distribution functions of switching points on the multiple price list, separately for men and women, and by task framing and perception. As in the main analyses, there are no big differences in switching behavior between framings for male and female participants. Only men who do not perceive the task in the stereotypical way (“Non-Stereotypical” subsample) switch at earlier points on the price list when being assigned to the male framing. Hence, these men seem to display a distaste of the matrix task in the male framing relative to the female framing. A tendency for such preferences has already been apparent in the main sample but is clearly stronger in the subsample considered here. The fact that this subsample comprises individuals that do *not* perceive the tasks in the intended way in terms of task stereotype suggests, however, that it

Table 13: Performance Beliefs by Framing (Stereotypical Subsample)

	(1)	(2)	(3)	(4)
	Baseline Performance Beliefs		Future Performance Beliefs	
	Men	Women	Men	Women
Male Framing	0.0546 (0.199)	-0.281* (0.155)	0.507*** (0.147)	-0.534*** (0.126)
Baseline Performance	0.377*** (0.0577)	0.538*** (0.0536)	0.644*** (0.0500)	0.723*** (0.0541)
Constant	0.350 (0.612)	2.209*** (0.560)	0.0632 (0.494)	1.844*** (0.455)
Controls	X	X	X	X
Observations	279	306	279	306
R-squared	0.216	0.297	0.511	0.524
Mean of Depend. Var.	3.953	3.418	4.731	4.278

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes: This table presents the results of regressions of baseline performance beliefs (columns (1) and (2)) and future performance beliefs (columns (3) and (4)) on the treatment indicator *Male Framing* and baseline performance by gender and for individuals in the “Stereotypical” subsample, i.e., participants whose stated task evaluations correspond to the intended task stereotype. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 3 observations that have non-valid values for ethnicity.

is not the gender congruency that leads men to move away earlier from the male-framed task. An alternative explanation for this behavior could be stated task enjoyment: If men report to enjoy working on the male-framed task less than working on the female-framed task, this could explain their earlier switching towards the outside option task when assigned to the male framing. However, as discussed in section 3.3.1, there are no significant differences in task enjoyment measures for men by task framing. Understanding the observed differences in switching behavior thus remains a task for future research.

Figures 15 and 16 now show switching along the multiple price list for the subsamples “Non-Stereotypical” and “Stereotypical” taking into account participants’ stated beliefs about their future task performance. Differences by task framing remain small, as observed in the main sample already. Only the right panel of Figure 16 suggests that women who perceive the tasks in the intended, stereotypical way switch away from the matrix task at later points in the male framing relative to their optimal switching point for maximizing expected earnings in work round 2. Given that this

Table 14: Task Enjoyment by Framing (Non-Stereotypical Subsample)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline Task Enjoyment		Expected Future Task Enjoyment		Task Enjoyment	
	Men	Women	Men	Women	Men	Women
Male Framing	-0.150 (0.234)	0.318 (0.229)	-0.282 (0.231)	0.0900 (0.226)	-0.170 (0.156)	-0.153 (0.157)
Past Perf. Beliefs	0.474*** (0.0887)	0.579*** (0.110)	0.274*** (0.0874)	0.313*** (0.112)	-0.0829 (0.0610)	-0.130* (0.0786)
Future Perf. Beliefs	0.383*** (0.105)	0.442*** (0.143)	0.672*** (0.0997)	0.914*** (0.134)	0.383*** (0.0717)	0.576*** (0.0900)
Baseline Performance	-0.0372 (0.0877)	0.0281 (0.136)	-0.0872 (0.0826)	-0.129 (0.128)	-0.0592 (0.0596)	-0.150** (0.0714)
Baseline Enjoyment					0.753*** (0.0394)	0.764*** (0.0377)
Constant	1.945 (1.678)	-0.962 (1.538)	1.635 (2.036)	-1.049 (1.673)	0.169 (0.864)	-0.314 (0.732)
Controls	X	X	X	X	X	X
Observations	414	393	414	393	414	393
R-squared	0.263	0.322	0.317	0.355	0.701	0.732
Mean of Depend. Var.	5.710	4.969	5.727	4.982	5.727	4.982

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table presents the results of regressions of baseline task enjoyment (columns (1) and (2)) and expected future task enjoyment (columns (3), (4), (5) and (6)) on the treatment indicator *Male Framing*, baseline and future performance beliefs and actual baseline performance by gender and for individuals in the “Non-Stereotypical” subsample, i.e., participants whose stated task evaluations do not correspond to the intended task stereotype. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). Columns (5) and (6) additionally control for reported baseline enjoyment. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 1 observation that prefers not to state their ethnicity in the survey.

pattern, already suggested in the main sample, becomes stronger in the “Stereotypical” subsample, it seems indeed to be that case the male stereotype drives this revealed preference displayed by women for the male-framed matrix task. While this finding goes against the initial hypothesis on taste-based self-stereotyping, it is an interesting result that deserves further investigation in future research. One potential explanation could be that the male stereotype provokes a defiant response of female experimental subjects, similar to experimenter demand but with the opposite intention. Such an explanation is related to the concept of psychological reactance discussed in the psychology literature in more detail (e.g., see [Rosenberg and Siegel 2018](#)). The fact that women for decades have

Table 15: Task Enjoyment by Framing (Stereotypical Subsample)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Task Enjoyment	Expected	Future	Task Enjoyment	
	Men	Women	Men	Women	Men	Women
Male Framing	0.121 (0.298)	-0.476* (0.277)	0.0911 (0.293)	-0.725** (0.286)	0.00594 (0.198)	-0.342* (0.179)
Past Perf. Beliefs	0.548*** (0.110)	0.693*** (0.127)	0.0740 (0.113)	0.281** (0.124)	-0.313*** (0.0885)	-0.277*** (0.0912)
Future Perf. Beliefs	0.592*** (0.133)	0.312** (0.154)	1.053*** (0.123)	0.753*** (0.163)	0.635*** (0.104)	0.501*** (0.121)
Baseline Performance	-0.347*** (0.110)	0.0290 (0.132)	-0.410*** (0.107)	-0.0529 (0.132)	-0.165** (0.0666)	-0.0762 (0.0975)
Baseline Enjoyment					0.706*** (0.0406)	0.806*** (0.0443)
Constant	1.683* (0.856)	2.310** (1.035)	1.310 (1.006)	2.382** (1.080)	0.122 (0.798)	0.521 (0.609)
Controls	X	X	X	X	X	X
Observations	279	306	279	306	279	306
R-squared	0.333	0.333	0.369	0.348	0.721	0.743
Mean of Depend. Var.	5.642	4.833	5.731	4.690	5.731	4.690

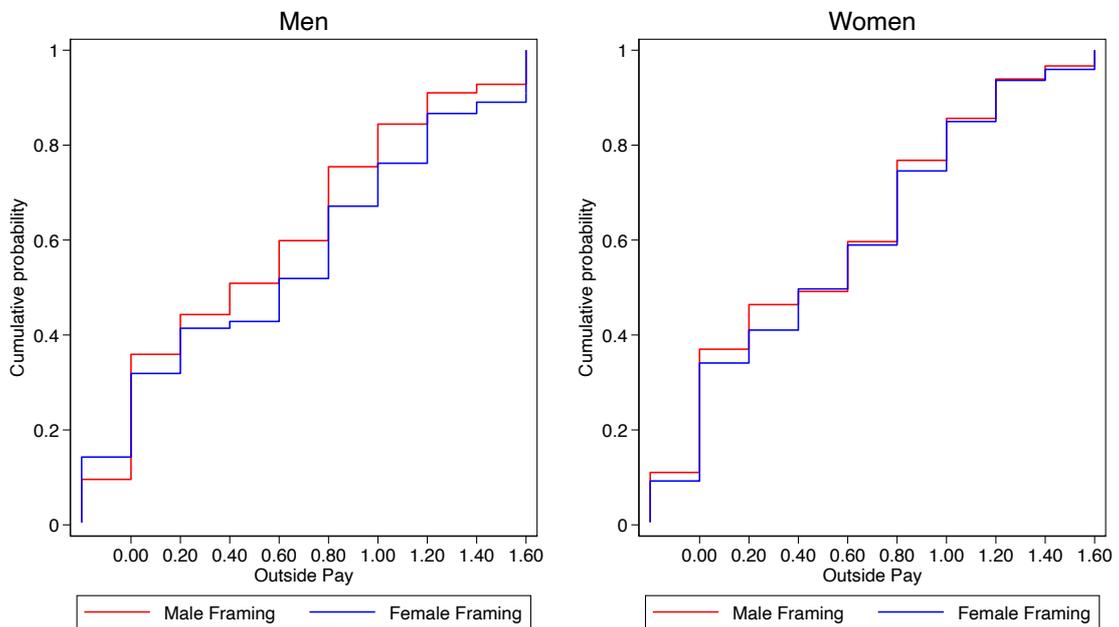
Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes: This table presents the results of regressions of baseline task enjoyment (columns (1) and (2)) and expected future task enjoyment (columns (3), (4), (5) and (6)) on the treatment indicator *Male Framing*, baseline and future performance beliefs and actual baseline performance by gender and for individuals in the “Stereotypical” subsample, i.e., participants whose stated task evaluations correspond to the intended task stereotype. *Male Framing* equals 1 if a participant saw the male framing of the matrix task, and 0 otherwise (i.e., if they saw the female framing). Columns (5) and (6) additionally control for reported baseline enjoyment. All columns control for age, ethnicity and education as categorical variables. Age has yearly categories from 18 to 25 years, ethnicity comprises the categories Black, White, Asian, Mixed and Other. Education has the categories less than a high-school diploma, high-school diploma & no college, high-school diploma & currently in college, associate degree, bachelors degree, graduate degree and higher, and other. I exclude 3 observations that have non-valid values for ethnicity.

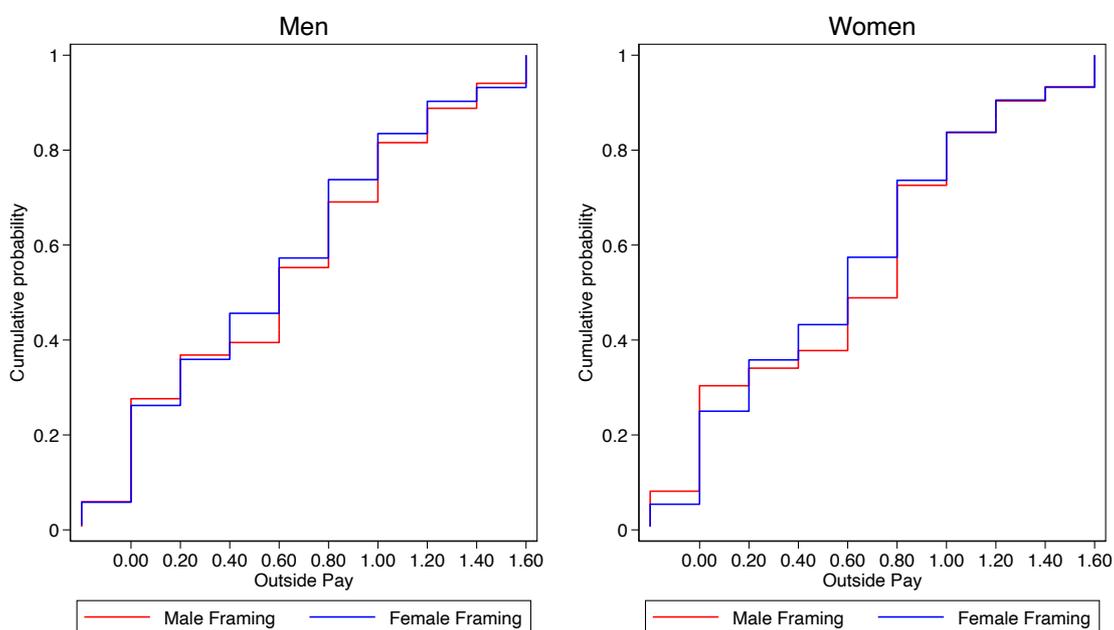
experienced exclusion from male-typed occupations, tasks and life domains could provide a plausible explanation for why such a response is only observable for female subjects. Of course, this is just a hypothesis consistent with the findings of the current study and rigorous further investigation is required to investigate this claim further.

Figure 13: CDF of Switching along MPL; Non-Stereotypical Task Perception



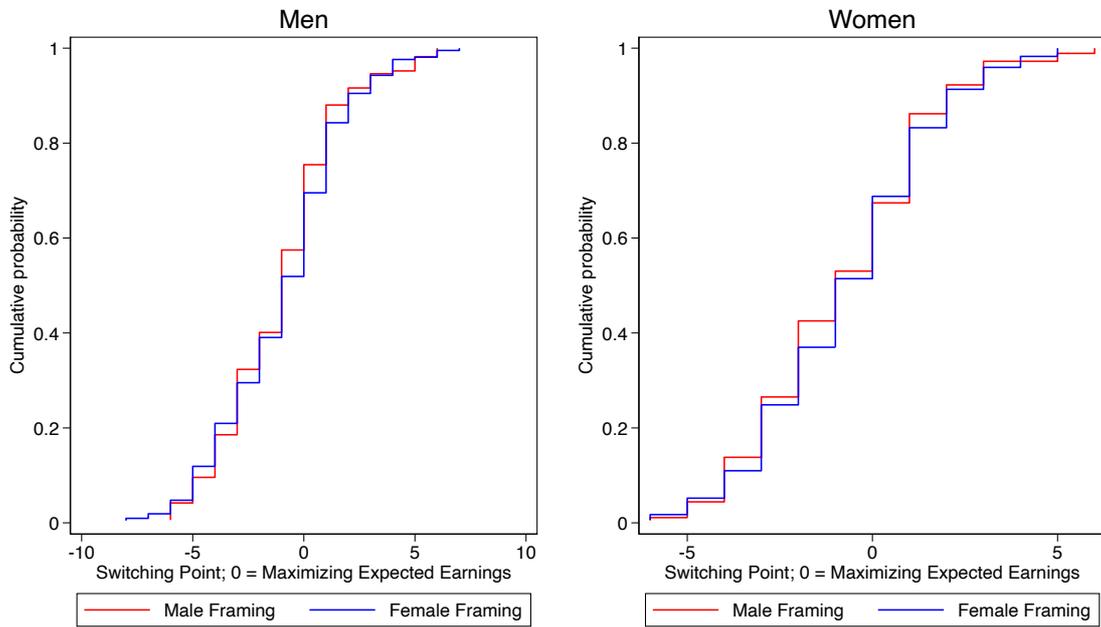
Notes: This figure shows the cumulative distribution functions of the switching points along the multiple price list, separately for men and women, by assigned task framing and for individuals in the “Non-Stereotypical” subsample, i.e., participants whose stated task evaluations do not correspond to the intended task stereotype. The cumulative probability is depicted on the y-axis, the fixed payment for the outside option as it increases along the MPL is shown on the x-axis. The figure excludes 77 observations with non-unique switching points along the MPL.

Figure 14: CDF of Switching along MPL; Stereotypical Task Perception



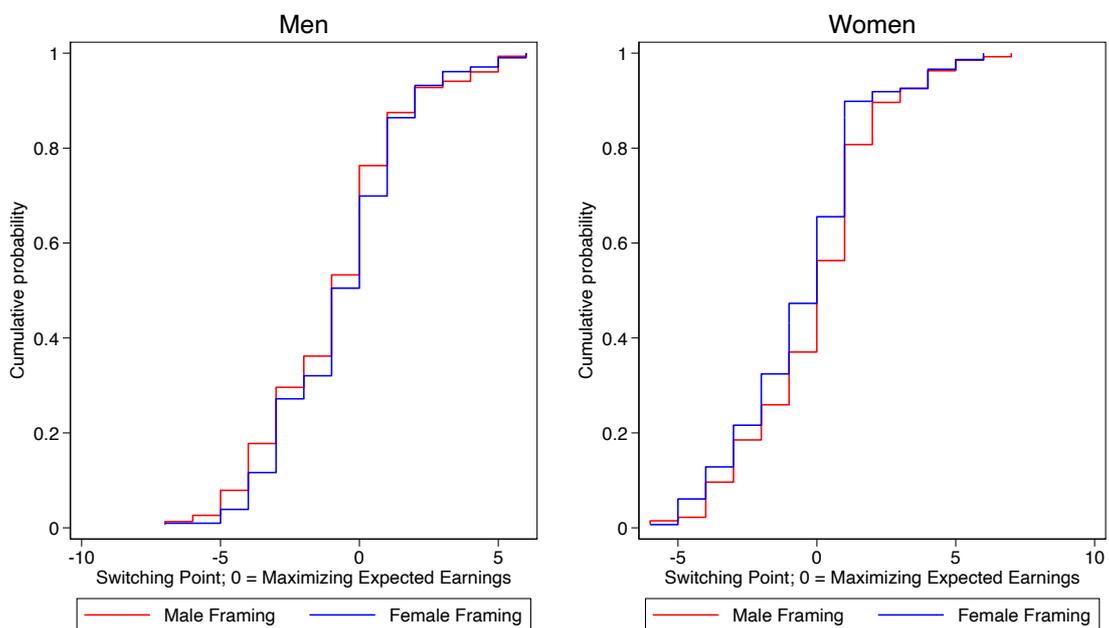
Notes: This figure shows the cumulative distribution functions of the switching points along the multiple price list, separately for men and women, by assigned task framing and for individuals in the “Stereotypical” subsample, i.e., participants whose stated task evaluations correspond to the intended task stereotype. The cumulative probability is depicted on the y-axis, the fixed payment for the outside option as it increases along the MPL is shown on the x-axis. The figure excludes 50 observations with non-unique switching points along the MPL.

Figure 15: CDF of Switching along MPL relative to Beliefs; Non-Stereotypical Task Perception



Notes: This figure shows the cumulative distribution functions of the switching points along the multiple price list relative to the switching point that would maximize expected earnings in work round 2 according to participants stated beliefs about their future performance, separately for men and women, by assigned task framing and for individuals in the “Non-Stereotypical” subsample, i.e., participants whose stated task evaluations do not correspond to the intended task stereotype. The cumulative probability is depicted on the y-axis, the switching point is shown on the x-axis. Here, the value of 0 describes the point at which participants should switch if they were maximizing expected earnings; switching points > 0 describe switching points after this point where participants express a preferences for the matrix task; switching points < 0 describe switching points before this point where participants express a distaste for the matrix task. The figure excludes 77 observations with non-unique switching points along the MPL.

Figure 16: CDF of Switching along MPL relative to Beliefs; Stereotypical Task Perception



Notes: This figure shows the cumulative distribution functions of the switching points along the multiple price list relative to the switching point that would maximize expected earnings in work round 2 according to participants stated beliefs about their future performance, separately for men and women, by assigned task framing and for individuals in the “Stereotypical” subsample, i.e., participants whose stated task evaluations correspond to the intended task stereotype. The cumulative probability is depicted on the y-axis, the switching point is shown on the x-axis. Here, the value of 0 describes the point at which participants should switch if they were maximizing expected earnings; switching points > 0 describe switching points after this point where participants express a preferences for the matrix task; switching points < 0 describe switching points before this point where participants express a distaste for the matrix task. The figure excludes 50 observations with non-unique switching points along the MPL.