# 1 Chapter 2: INSERT TITLE

## 1.1 Introduction

## 1.2 Rising use of performance pay, especially competitive pay over time - and how it may contribute to the gender wage gap/other gender diff in labor market outcomes

Compensation packages based on performance pay, such as bonuses, commissions, and piece-rate payments, have risen in popularity relative to hourly/salaried pay, especially among workers in the highest tiers of occupations [@Hall1998; @Murphy1999; @Cunat2005; @Lemiuex2009]. There is evidence that the increasing use of performance pay lends itself to wage inequality. @Lemiuex2009 showed that an increased dependence on performance pay during the late 1970’s and early 1990’s accounted for 21% of the observed growth in variance of male wages. Bonuses and commissions, arguably the most competitive compensation schemes, may be especially important in driving the large disparity between the highest and lowest percentile earners within organizations [@Bell2010; @Bell2014; @Benabou2016]. Importantly, performance pay may contribute to the gender wage gap too. Using data from the National Longitudinal Surveys of Youth, @McGee2015 show that women are less likely to be employed in occupations that receive bonuses, and simultaneously are more likely to receive piece-rate pay – the least competitive of all forms of performance pay, where workers are paid based on their absolute output.

## 1.3 Gender differences in competitiveness as possible contributor to labor market gaps

Since competition is relevant to labor market outcomes, researchers began to focus on how a person’s gender affects their competitiveness, both in terms of willingness to enter competitions and response when required to enter a competition. To date, most of the research on gender differences in competitiveness has focused on either i) explaining the sources of the gender difference [e.g., @Veldhuizen2017] or ii) designing interventions to encourage women to compete more [@Balafoutas2012; @Sutter2016; @Cassar2016; @Brandts2015; @Niederle2013; @Brandts2015; @Healy2011; @Alan2018]. Less consideration has been paid to how competitions may differentially, and perhaps negatively, impact women.

The introduction of Chapter 1 provides an overview of the literature on gender differences in willingness to compete, so we will only review the literature on gender differences in response to entering competitive environments here. There are three major time points at which competition may affect men and women differently: before, during, and after competition. The majority of previous studies in this space have examined gender differences in response to competition during and after performance, which we will briefly review here.

## 1.4 Gender differences in response to competitive environments

### 1.4.1 During competition

Gender differences in performance during competition: Although competitions are generally motivating and designed to improve performance by increasing effort [@Connelly2014a; @Murayama2012; @Miller2019a], previous research suggests that men perform better under competitive payment schemes relative to non-competitive payment schemes, while women’s performance does not respond to competitions [@Gneezy2003; @Gneezy2004; @Gunther2010; @Samak2013]. @Gneezy2003 show that there is no gender difference in performance when participants are solving mazes following a piece-rate payment scheme, but a significant gender difference in performance arises under a tournament payment scheme, with males performing better. @Gunther2010 replicate the effect of competition on gender differences in performance for a male-typed task, but find no gender differences in performance during competition for female-typed or gender-neutral tasks.

INSERT OTHER CITES:

* Evidence that women’s performance doesn’t strongly respond to competition compared to men: <https://onlinelibrary.wiley.com/doi/full/10.1111/ecca.12417>
* Suggests that women may not respond well to competitive pressure (aka when stress is kept to a minimum, there are no gender differences in performance, but when certain knock-out rules are applied, a difference emerges): <https://www.sciencedirect.com/science/article/pii/S0167268121001785?casa_token=1G3VrTCCNu8AAAAA:dsOsjejPKHnunOTRSqkEHU-odJMjDPhHUBXy-dTr9_JPX4KqAqrH4bihs5riR7gypyza2Rko_vg>
* there is a growing literature showing that women are less willing to guess on exams [@Pekkarinen2015; @Baldiga2014; @Iriberri2021], which in turn negatively impacts performance on said exams [@Pekkarinen2015; @Baldiga2014] - which they argue may driven by women being less confident in their probability of answering correctly or being more risk averse. @Riener2018a suggests this phenomenon starts at an early age, with girls as young as 8 years of age being significantly less willing to guess on exams relative to men
* @Paserman2007: “Data on serve speed, on first serve percentages and on rally length suggest that women play a more conservative and less aggressive strategy as points become more important.”

### 1.4.2 After competition

Gender differences in response to losing: During repeated competition, women tend to perform worse in subsequent performance rounds after losing, even if the monetary prize they lost was relatively meager, while men only perform worse in subsequent rounds if they lost the chance to win a large monetary prize [@Gill2014]. Other research suggests women stop competing altogether after losing if given the choice. @Buser2019, who examine the effects of losing while competing in the Dutch Math Olympiad on the choice to compete in subsequent years, show that men are just as likely to compete even if they lost the previous year, while women are less likely to compete again if they lost before. Overall, this body of literature suggests that competitions may differentially impact women and men, both during and after the competition.

INSERT OTHER CITES:

* negative feedback increases women’s likelihood of dropping out of their major: @Astorne-Figari2018
* <https://drive.google.com/file/d/1eMZJpkqa0QvDhcf76r2U8bkuelVa1Byt/view>: “We find that, among assistant professors, a flat rejection reduces the confidence in publishing the paper in any leading journal to a significantly greater extent for women than it does for men. We find no gender differences among associate and full professors, likely due to survivorship bias.”
* <https://www.nber.org/system/files/working_papers/w29382/w29382.pdf>: “We find that, holding fixed performance and decisions before feedback, women update their beliefs and choices more negatively than men do after bad news.”

### 1.4.3 Before competition

As mentioned previously, little research has examined how competitions may affect gender differences in behavior during arguably the most critical period: before an individual enters a competition, where they have the most control of their subsequent performance in the competition. Given previous research suggesting that women and men may respond differently during and after competitions, we expect that they will also employ different behaviors and have different perceptions of themselves and others in advance of a competition.

We only know of a few studies that explore this open question: insert possible cites if relevant

#### 1.4.3.1 Preparation as a coping strategy before competition & possible mechanisms

Preparing for a competition, through either practicing or studying, is a coping strategy individuals may employ before entering a competition. Since competitions, by definition, compare the performance among two or more individuals, they naturally lead to self-evaluation and comparative judgments of self with others - processes that are intimately linked to confidence. To the extent that confidence influences how much individuals think they need to prepare in order to win, we may expect to see women preparing more than men, particularly in competitive contexts, which naturally invoke self-other assessments. Thus, less confident individuals may prepare more. Moreover, they may prepare more in order to reduce the negative feelings caused by low confidence independent of any ambitions to win, since mastery is an important driver of confidence [@Gist1992; @Usher2008]. There is no theoretical or empirical reason to suspect that women would be less concerned with mastery than men. In fact, research suggests that women are just as likely as men to compete when competing against their own past performance, suggesting, at minimum, an equal desire for self-improvement [@Apicella2017a]. Similarly, given the inherent risk of competitive payment schemes relative to non-competitive payment schemes, it is possible that the aforementioned gender differences in risk attitudes may also lead women to be more likely to cope by preparing before performing in a competition relative to men.

#### 1.4.3.2 Gender stereotypes as a possible mechanism

A novel prediction deriving from the results showing robust perceptions of gender differences in preparation across all studies in Chapter 1 is that gender differences in preparing may be driven by persistent stereotypes of men and womens’ tendencies to prepare before performance.

##### 1.4.3.2.1 Prominence and characteristics of gender stereotypes

* Gender stereotypes derive from observers’ automatic tendency to make correspondent inferences about men and women’s dispositions [@Gilbert1995; @Ross1977; @Jones1967; @Gawronski2004]. These correspondent inferences have led to prominent gender stereotypes that exist across cultures [@Williams1990; @Williams1982; @Steinmetz2014; @Fiske2017].
* Stereotypes involve prescriptive, proscriptive, and descriptive components [@Prentice2002], where prescriptive and proscriptive stereotypes reflect cognitive representations of the characteristics women and men should and should not have, respectively, while descriptive stereotypes are representations of the typical man and woman [@Burgess1999].
* Gender stereotypes encompass a variety of attributes. For instance, there are physical (e.g., women are dainty), cognitive (e.g., men are analytical), and personality-based (e.g., women are nurturing) stereotypes [@Cejka1999; @Deaux1984].

##### 1.4.3.2.2 Implications of gender stereotypes for behavior

* Importantly, there is evidence that gender stereotypes can affect behavior [INSERT cites]. For instance, @Coffman2014a show that both men and women are less likely to contribute ideas to a group decision in gender-incongruent decision-making domains (e.g., women contributing ideas to a decision in the domain of sports), even when the group would have made a better decision with their contribution. [INSERT other example showing stereotypes affecting behavior].

Given the extensive evidence that gender stereotypes affect subsequent behavior, we expect that our findings of robust perceptions of gender differences in preparation likely contribute to gender differences in actual preparation behavior.

## 1.5 The current experiment

Overall, women may engage in more coping strategies than men, such as preparation, before entering competitions because they tend to be more risk-averse [@Croson2009; @Dohmen2011b; @Eckel2008; @Bertrand2010a] and less confident [@Bertrand2010; @Lundeberg1994; @Mobius2011; @Barber2001; @Croson2009], and/or may be adhering to gender stereotypes [insert coffman etc cites]. In support of this possibility, in Chapter 1 of this dissertation, we found evidence of a sizable gender difference in preparation, where women were more likely than men to choose to prepare before completing a multiplication task.

Here, we focus on how women and men differentially respond to competition through preparation. We expect to see both gender differences in actual preparation behavior, along with gender differences in perceptions of relative preparation, especially when men and women are required to compete (relative to non-competitive environments). More specifically, in the study included in this chapter, we tested whether competition exacerbates previously established gender differences in preparation by manipulating participants’ assigned payment scheme (i.e., competitive tournament payment scheme or non-competitive piece-rate payment scheme). We hypothesized that women will choose to practice problems at a higher rate than men, especially when assigned to the competitive tournament payment scheme (i.e., we anticipate a main effect of gender on the choice to practice, and an interaction between gender and condition, such that women will practice more than men in both conditions, but the difference-in-differences between practicing rates across genders will be greater in the competition condition). To be clear, there was no interaction between gender and choice to compete on the choice to prepare in any of the previous studies from Chapter 1. That is, women prepared more than men regardless of which payment scheme they had chosen. However, it is not possible to draw conclusions from this because i) we did not manipulate the payment scheme, so there could have been selection effects on one’s choice to prepare across payment schemes, such that those who were more likely to choose to compete may have been less likely to prepare, and ii) there was little power to detect any possible interaction effects. For instance, an average of only INSERT% (*N* = INSERT) of all women across the three studies in Chapter 1 chose to compete. Through the proposed experiment, we intend to expand upon the studies in Chapter 1 by directly manipulating participants’ payment scheme and recruiting a large sample to provide power to detect small effects.

We also tested whether gender predicts participants’ perceptions of their relative amount of preparation, given our hypothesis based on Study 3 of Chapter 1 that women may be especially susceptible to feelings of underpreparation relative to others when they have unlimited time to prepare. More concretely, we expected women will be more likely to assume they practice less than others compared to men (that is, the effect of gender on perceptions of relative practice will be negative), especially when assigned to the competitive tournament payment scheme (such that women in general will think that they practice less than other participants than men, but this difference will be exacerbated in the competition condition).

The research design, hypotheses, measures and analyses for this chapter were pre-registered on [OSF](https://osf.io/8bwfz/) and all analyses were conducted in R statistical software (version 4.0.4).

### 1.5.1 Methods

## 1.6 Participants

Participants were recruited on Amazon Mechanical Turk using the same screening criteria as all previous studies in Chapter 1. Like the last study of Chapter 1, we used Qualtrics’ fraud detection software to filter out responses that were suspicious either because they were likely 1) bots and/or 2) duplicate responses using the same exclusion criteria from before. These exclusions were applied for all main analyses reported in the results section.

The final dataset consists of 3980 participants (57.36% women), with an average age of 41.3 (*SD* = 13.2) years. Of the final sample, 75 participants (30.67% women) dropped out of the study before finishing and 192 participants were flagged by Qualtrics’ fraud detection software as suspicious based on the aforementioned criteria. We include analyses for the full sample in the appendix and all results are unchanged (INSERT DOUBLE CHECK).

## 1.7 Procedures

Participants included in the study were told they would be completing a multiplication task. Notably, we aimed to recruit a larger sample to grant us enough power to detect our anticipated interaction effects, and decided to shorten the task from two minutes to 30 seconds. Otherwise, the task used was identical to the ones used in previous studies.

Like the studies in Chapter 1, participants were first told about the rules for the multiplication task and were required to pass the same comprehension questions used in the previous studies before moving onto the main manipulation of payment scheme.

### 1.7.1 Manipulation of payment scheme

Unlike previous studies, participants were not able to choose a payment scheme. Instead, they were told about their random assignment to one of two payment schemes: the non-competitive piece-rate payment scheme or a competitive tournament payment scheme. Men and women were evenly assigned to both conditions. If they were assigned to the piece-rate payment scheme, they were paid $.10 per problem solved correctly. If they were assigned to the tournament payment scheme, they were randomly matched with another participant that was also assigned to that payment scheme and received $.20 per problem if they solved more problems than the other participant. Otherwise, they received nothing.

Again, we checked that condition was assigned evenly across participants (control= 50.21%) and genders included in the study. Of the men who completed the study, 50.38% were assigned to the control condition and of the women who completed the study, 49.72% were assigned to the control condition. We also assessed condition-dependent attrition by identifying the number of participants that dropped out during/after learning about condition and find that a relatively small proportion of participants out of the total sample dropped out after learning about their respective condition (*N* = 42). Given the small sample that dropped out relative to the total number of participants in the study, we are not concerned that condition-dependent attrition is driving any of the effects found in this study.

### 1.7.2 Main dependent variables of interest: Measures of preparation and perceptions of relative preparation

After they were informed of their payment scheme, all participants were given the opportunity to spend unlimited time preparing before completing the paid multiplication task. The nature of the unlimited preparation was identical to that used in Study 3 of Chapter 1, where participants who chose to prepare were shown 10 multiplication problems that were created randomly by drawing from the pool of numbers used in the main multiplication task. Unlike the last study in Chapter 1, participants were not asked to explicitly indicate whether they would like to study the times tables. Instead, they were shown the times table right after the practice problems directly on the practice problems page and told they could check their answers using the table as desired. By including the option to check their answers, we hoped to make the practice itself more useful by providing participants a way to receive feedback on their responses. At the bottom of each practice page, participants were asked if they would like to continue practicing multiplication problems, with the option to continue as many times as desired or opt out at any point. The amount of time (in seconds) participants spent on each practice page was also recorded. Thus, like the previous studies, we have multiple measures of preparation by design: 1) the actual decision to practice problems (before knowing what the practice entails), 2) among participants who chose to practice problems, the number of practice problems participants attempted (quantified as number of practice problems not left blank, irrespective of accuracy), 3) among participants who chose to practice problems, the amount of time they spent across all practice rounds they completed, and 4) the number of extra practice rounds participants completed after having completed the first round of practice. Again, the number of extra practice rounds serves as a way to quantify the number of times participants continue to practice after having seen what the practicing/studying looks like and having gone through it at least once. By encoding participants who both chose not to practice and those who chose not to continue practicing after the first round with zeroes in the dataset when creating this variable, we are able to separate out the effect of the choice to practice from the choice to continue practicing.

After completing the practicing/studying, participants guessed how much their amount of practicing for the task compared to all other participants that completed the task by indicating the decile of their practice relative to other participants. We also asked participants to indicate their anticipated decile when their amount of practicing was compared to that of all participants who identified as men and women, respectively.

We used these the decile questions to create the perceived practice deviation variables as follows: self-rated decile (either based on the question about practicing relative to all other participants, relative to only men, or relative to only women) - actual percentile based on number of practice problems completed. Therefore, negative values for this variable indicate a participant expected to have practiced less, relative to other participants, than they actually did, and vice versa for positive values. A value of zero, therefore, indicates that a given participant was completely accurate in their guess of relative practicing.

### 1.7.3 Paid multiplication task and post-task measures

After practicing, participants completed the paid multiplication task, received feedback about their absolute (but not relative) performance, and completed many of the same follow-up questions used across Chapter 1, including risk attitudes, confidence, and perceptions of gender differences in preparation, competitiveness, and performance. One of the perceptions of gender differences questions deviated slightly from the previous studies, which was edited for the sake of clarity. Instead of asking participants to indicate “Do you think men or women in this study chose the tournament payment option more often?”, they were asked “If given the opportunity to choose between the two payment schemes (Piece Rate or Tournament), do you think men in this study would choose the piece rate or the tournament payment scheme more often?”, with the options to indicate: “Men would choose tournament more often than piece rate”, “Men would choose piece rate more often than tournament”, or “Men would choose each payment scheme equally”. This question was repeated with respect to women in the study.

We paid participants to answer the questions about their confidence and perceptions of gender differences correctly at the same rate as previous studies. Finally, they completed the same demographic questions from Chapter 1 and provided feedback on the study before being paid for their participation.

### 1.7.4 Results

## 1.8 Describing main variables of interest

First, we explored the characteristics of the main practice variables in the dataset. 45.51% of all participants chose to practice, with 48.22% choosing to practice in the piece-rate payment condition and 51.78% choosing to practice in the tournament payment condition. This difference in the choice to practice across conditions is significant when condition is included as a predictor alone, , 95% CI , , , , but in the subsequent section we explain how the effect changes when including other predictors in the model. Participants spent an average of 29.12 seconds practicing across all rounds of practice and of those who chose to practice, completed 0.14 total rounds of extra practice problems (that is, rounds of practice after having seen what the practice looks like).

Like all studies in the first chapter, we replicate the effect of gender on risk, , 95% CI , , , , and confidence, , 95% CI , , , , such that women were more risk averse and less confident relative to men.

Contrary to the majority of studies in the first chapter, we find a significant effect of gender on task score, Mwomen=10.45, SD=4.47; Mmen= 12.29, sd=7.28, even when including risk, confidence, and an interaction between gender and condition in the model, , 95% CI , , , . We explore this finding further in the discussion section for this study.

## 1.9 Effects of gender and condition on both practicing and perceptions of one’s relative practicing

Through a logistic regression with gender, condition, and the interaction between the two predicting the binary choice to practice problems, we replicate the effect of gender on the choice to practice found in Chapter 1, , 95% CI , , , , where 50.74% of women chose to prepare relative to 37.62% of men (see Figure 1.1). However, we do not find an interaction between gender and condition, , 95% CI , , , , contrary to our hypothesis that the gender difference in the choice to prepare would be exacerbated under the tournament payment scheme relative to the piece-rate payment scheme. Additionally, the aforementioned effect of condition on the choice to practice is no longer significant in the model including these additional predictors, , 95% CI , , , .

As part of our pre-registered analyses, we also explored other measures of practice to test the robustness of the effect of gender on practicing. We find that women, relative to men, completed a significantly higher number of practice problems, , 95% CI , , , , more rounds of extra practice , 95% CI , , , , and spent more time completing practice problems, , 95% CI , , , .

Based on previous literature on risk aversion and confidence affecting competition entry, we expected participants with especially high levels of risk aversion and/or low levels of confidence would be especially likely to choose to practice before entering a competition, and that this effect may interact with gender. Thus, we tested possible three-way interactions between gender, condition, and risk aversion or confidence on the choice to practice problems through two pre-registered logistic regressions, but did not find evidence that risk aversion, INSERT, nor confidence, INSERT, interacted with gender and condition.

## 1.10 Accuracy of levels of practicing based on participant gender

Next, we ran a linear regression with gender, condition, and the interaction between those two variables predicting the aforementioned perceived practice deviation variable (that is, subtracting each participants’ percentile based on number of practice problems completed from their self-rated decile) to test our second hypothesis that women would be more likely to assume they practice less than others compared to men, especially under the competitive tournament payment scheme. Though we did not find evidence of the anticipated interaction effect, , 95% CI , , , , we find a significant effect of gender on perceived practice deviation, such that women (relative to men) were significantly less likely to assume they practice more than others, , 95% CI , , , , Mwomen=23.56, SD=56.11; Mmen= 39.69, sd=54.87 (see Figure 1.2). We performed a more targeted exploratory analysis to see if the effect held when participants have actually practiced (and as a result, the question about their relative practicing may have felt more relevant), and find that among this subset of the data, women (again, relative to men) were still significantly less likely to believe that they practiced more than others, , 95% CI , , , .

Since this is the first time we have used the perceived practice deviation variable and are not able to attest to its robustness, we also explored another way of testing this hypothesized effect by using participants’ self-rated decile as the dependent variable instead of perceived practice deviation and then controlling for number of practice problems attempted (as a proxy for more precise estimate of amount of practicing) in a linear regression. We find that, regardless of the number of practice problems attempted, women are significantly less likely to say they practice more than others compared to men, INSERT, although this effect does not hold when focusing on the subset of participants who chose to practice, INSERT.

On top of the differences in how much women and men in this study perceived they practiced relative to others, we also pre-registered tests of men and women’s accuracy of their relative practice through a series of t-tests comparing the perceived practice deviation variable to 0 (which would represent a participant guessing their exact decile correctly). Across the full dataset, most participants tended to overestimate how much they practiced relative to others, INSERT. After honing in on each gender included in the study, we find that this effect holds among both women, INSERT, and men, INSERT. Notably, participants who chose to practice significantly underestimated their relative practice, both among women, INSERT, and men, INSERT.

We also explored how self-rated decile changes based on whether participants were asked to compare their amount of practicing to men or women in the study specifically, and find that participants’ perceptions of how much they practiced relative to women in the study are significantly lower than perceptions of much they practiced relative to men, , 95% CI , , , .

## 1.11 Perceptions of gender differences in behavior

Across all measures of perceptions of gender differences in behavior, we replicate effects found in the previous studies. First, the majority of participants (59.57%) said that women would be more likely to practice/study for the task, which was significantly higher than the proportion of participants that said men would be more likely to practice/study than women (4.73%) and the proportion of participants that said there was no difference in the likelihood that men and women would practice/study (35.7%), , .

Similarly, participants were significantly more likely to say that women prepare more than men in general (68.28% of participants), , , relative to the proportion of participants that said men prepare more than women (4.41% of participants) or that there is no difference in how much men and women prepare (27.31% of participants).

Yet, participants did not expect a gender difference in performance on the main multiplication task used, , , where 54.17% of participants said that there was no difference in how many multiplication problems men and women correctly solved, while 20.56% said men correctly solved more multiplication problems than women and 25.27% said women had a performance advantage over men.

Finally, 64.24% of participants expected women would be more likely to choose the piece-rate payment scheme than the tournament payment scheme, which was a significantly higher proportion of participants than those who expected women would choose each payment scheme equally (20.9%) and than those who expected women would choose tournament more often than piece rate, (14.86%), , . On the contrary, when asked about how much men in the study would compete, a significant majority of participants (63.5%) expected men to be more likely to choose the tournament payment scheme over the piece-rate payment scheme, relative to the proportion of participants who said men would choose each payment scheme equally (15.8%) and the proportion who said men would choose piece rate more often than tournament (20.7%), , .

## 1.12 Figures

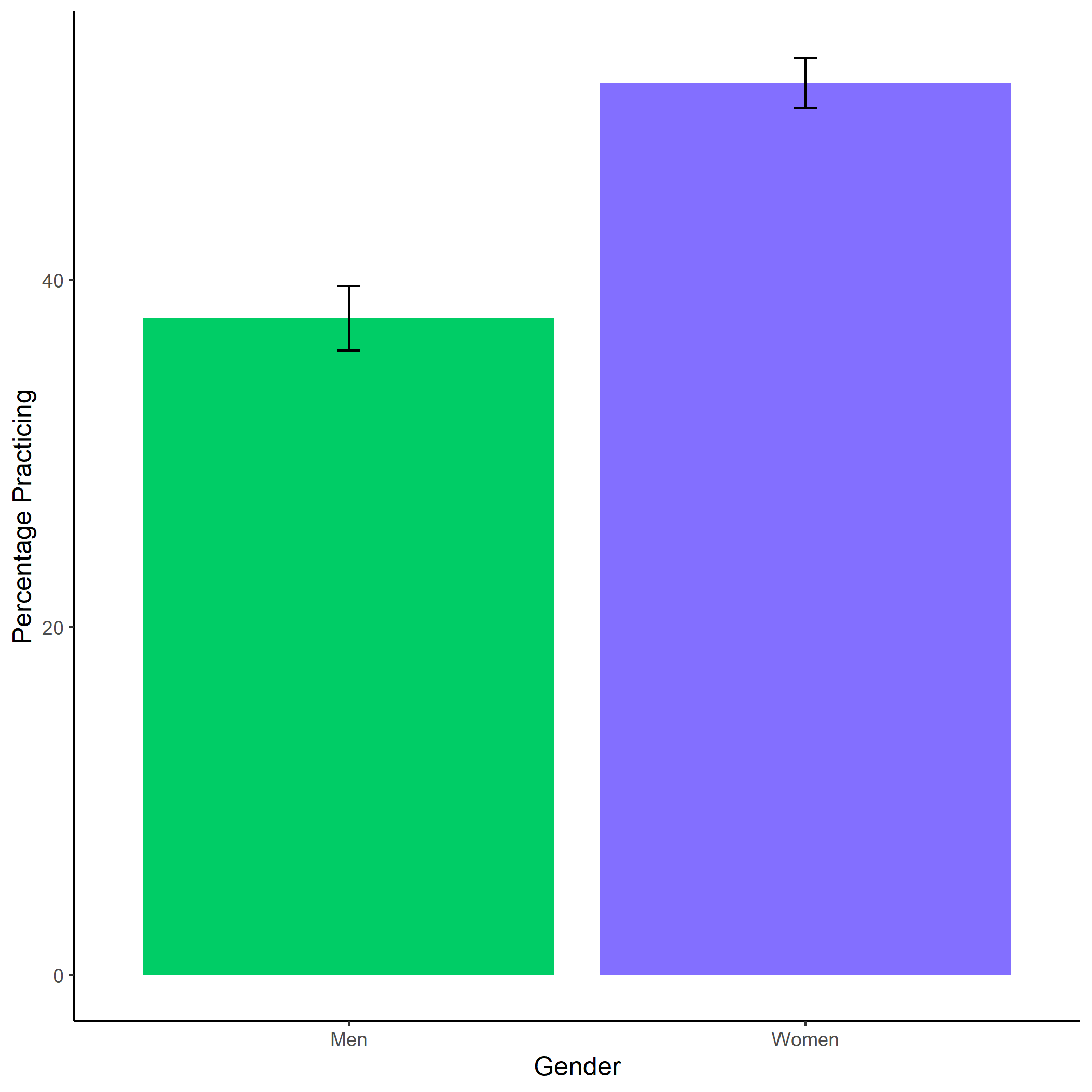


Figure 1.1: Proportion of women (INSERT%) and men (INSERT%) who chose to prepare by condition. Error bars represent standard errors - insert denominators for SE bars.

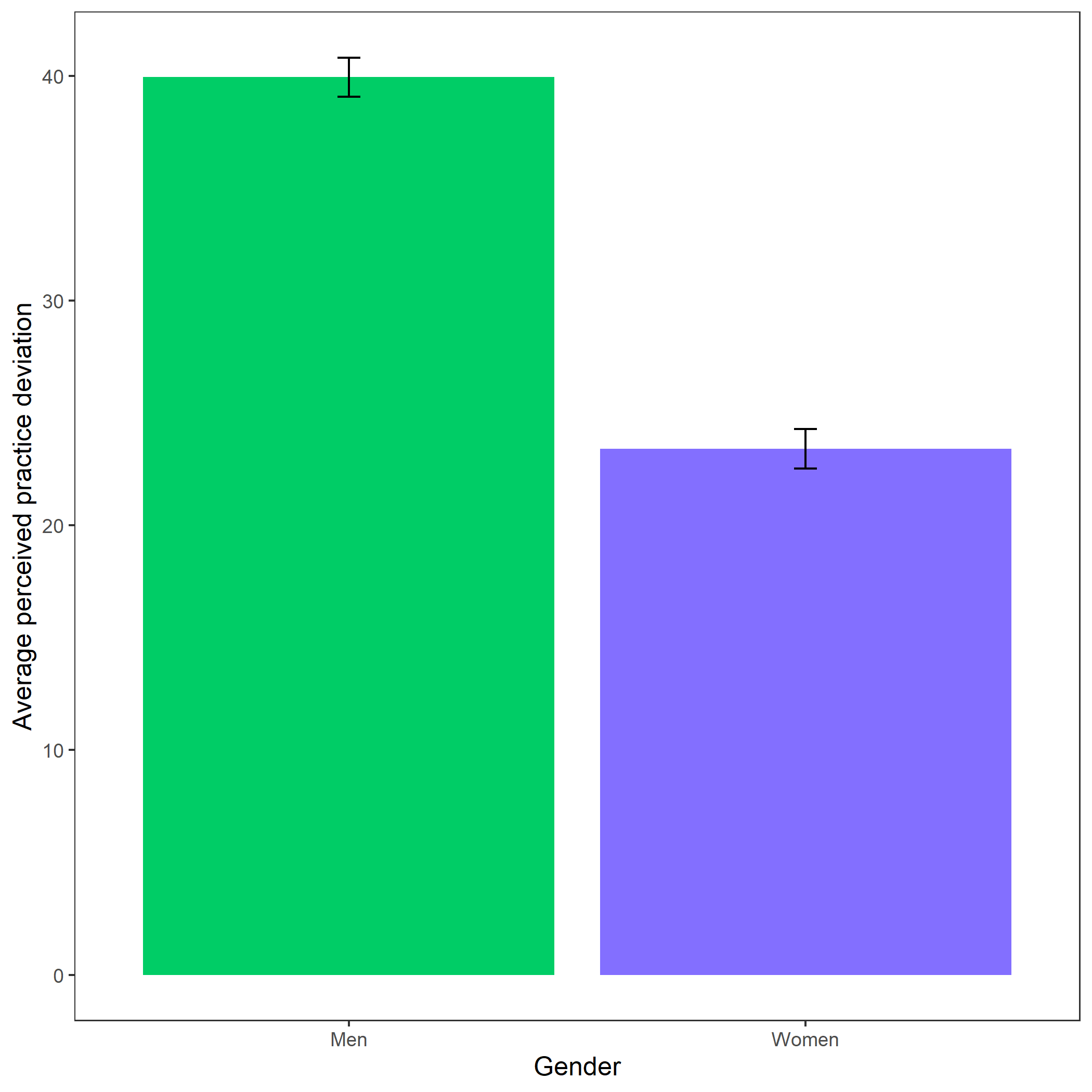


Figure 1.2: Perceived practice deviation based on participant gender. Error bars represent standard errors - insert denominators for SE bars..