



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Does Increasing the Riskiness of Choices Widen Gender Gaps?

Samuel D. Hirshman, Alexander Willén

To cite this article:

Samuel D. Hirshman, Alexander Willén (2025) Does Increasing the Riskiness of Choices Widen Gender Gaps?.
Management Science 71(8):6880–6901. <https://doi.org/10.1287/mnsc.2023.04088>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. You are free to download this work and share with others for any purpose, except commercially, and you must attribute this work as “*Management Science*. Copyright © 2024 The Author(s). <https://doi.org/10.1287/mnsc.2023.04088>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc/4.0/>.”

Copyright © 2024 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Does Increasing the Riskiness of Choices Widen Gender Gaps?

Samuel D. Hirshman,^{a,*} Alexander Willén^a

^aDepartment of Economics, Norwegian School of Economics, Bergen 5045, Norway

*Corresponding author

Contact: samuel.hirshman@nhh.no,  <https://orcid.org/0000-0002-0330-9375> (SDH); alexander.willen@nhh.no (AW)

Received: December 13, 2023

Revised: April 29, 2024

Accepted: August 4, 2024

Published Online in Articles in Advance:
November 22, 2024

<https://doi.org/10.1287/mnsc.2023.04088>

Copyright: © 2024 The Author(s)

Abstract. We isolate the causal effect of changing the riskiness of choices on the gender gap in real world outcomes. We do so by exploiting a national reform to the regrade system of Norwegian universities that generated exogenous variation in the probabilities of the outcome of regrade requests. Using unique individual-level administrative data, we show that both the expected value of a regrade request, as well as the downside risk, increased substantially as a result of the policy. We then show how the ostensibly gender-neutral policy substantially increased gaps between men and women because they differed in their risk preferences and beliefs. Specifically, the exogenous change in the riskiness of requesting a regrade augmented the gender gap in regrade requests by 90%. We demonstrate that this has important implications for students through its impact on their grades. We disentangle the relative importance of mechanisms through auxiliary reduced-form analyses, structural estimation, and a supplemental laboratory experiment. Taken together, our mechanism analyses suggest that gender differences in risk preferences and loss aversion are not sufficient to explain the increased gender gap after the policy change, and that gender differences in confidence or beliefs are necessary to rationalize the data and more broadly consistent with the patterns we observe.

History: Accepted by Dorothea Kubler, behavioral economics and decision analysis.



Open Access Statement: This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. You are free to download this work and share with others for any purpose, except commercially, and you must attribute this work as “*Management Science*. Copyright © 2024 The Author(s). <https://doi.org/10.1287/mnsc.2023.04088>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc/4.0/>.”

Funding: This project was partially funded by the Research Council of Norway through its Centers of Excellence Scheme [FAIR Project 262675].

Supplemental Material: The online appendix and data files are available at <https://doi.org/10.1287/mnsc.2023.04088>.

Keywords: risk • gender gaps • educational choices • inequality

1. Introduction

Many policy reforms affect the riskiness of choices. This is true across a broad range of areas: education finance reforms change the riskiness of taking student loans, crime policies alter the riskiness of breaking the law, changes to the welfare state affect the riskiness of starting new businesses, and changes to worker incentive and bonus schemes influence the riskiness of the specific work strategies and decisions that individuals take in their professional lives. However, if there are systematic differences in risk preferences and beliefs across demographic groups, then these groups may respond very differently to changes in the riskiness of choices. Consequently, even ostensibly neutral and nondiscriminatory policies may generate differences in program take-up and exacerbate inequalities across groups.

Two demographic groups that may respond very differentially to changes in the riskiness of choices are

men and women. A substantial body of research in the social sciences has documented gender differences in a variety of behaviors and preferences. For example, women tend to be less confident (Barber and Odean 2001, Niederle and Vesterlund 2007, Exley and Kessler 2022, Exley and Nielsen 2024), more likely to opt out of competition (Gneezy et al. 2003, Niederle and Vesterlund 2007, Balafoutas and Sutter 2012, Sutter and Glätzle-Rützler 2015), and more risk averse in certain contexts (Croson and Gneezy 2009, von Gaudecker et al. 2011, Charness and Gneezy 2012, Niederle 2015, Filippin and Crosetto 2016, Buser et al. 2021, Thöni and Volk 2021). Prior work also suggests that gender differences in primitives like preferences, beliefs, and confidence impact gender gaps in labor market outcomes. However, field demonstrations have focused on correlations between choices and laboratory-elicited preferences (Buser et al. 2014, Cortés et al. 2023, Li and Zafar 2023). Outside the laboratory, there is very little causal

work on the consequences of gender differences in beliefs and risk preferences (Flory et al. 2015, Samek 2019). Estimating causal effects of these primitives presents a challenging identification problem. Specifically, tying gender gaps in real world outcomes to specific behavioral mechanisms not only requires detailed individual-level data, but also multiple sources of exogenous variation (one that allows identification of the reduced-form effects and one that identifies how a particular character trait interacts with that policy) or a narrowly focused policy change.

We take advantage of rich administrative data and combine that with policy-induced exogenous variation in the riskiness of choices to overcome these challenges and isolate the causal effect of changing the riskiness of choices—altering the probabilities of the outcomes associated with choices—on the gender gap in real world outcomes. The national policy we study was implemented in the second half of 2014 and altered the grading procedure for Norwegian university students who requested regrades on assignments and exams. In Norway, all students have the legal right to request a regrade of any assignment that impacts their final grade. The regrade is performed by a new examiner who was not involved in the original grade decision, and the decision of that examiner is final. Prior to the reform, the new examiner would receive the assignment and would be informed about the student's initial grade and the original examiner's explanation of that initial grade. After the reform, the new examiner received only the student's assignment. The goal of the reform was to provide students with a fair chance at reassessment not influenced by the prior grade result.

By exploiting unique individual-level administrative data from Norway's leading business school on all student assignments, grades, and regrades, including both requests and outcomes, we document a substantial increase in the likelihood that a grade gets changed as a result of the policy. Before the policy was implemented, 84.4% of regrade requests resulted in an unchanged grade, 14.7% resulted in a grade improvement, and 0.9% led to a grade reduction. Following the reform, the expected value of a regrade request increased, but so did the downside risk. Specifically, after the reform, the share of regrades that receive a negative reevaluation increases to 11.7% and the share of regrades that receive a favorable reevaluation increases to 32.2%.¹ Because the reform narrowly impacts the probabilities associated with the different possible regrade outcomes while leaving all other elements of the regrade process unchanged, it provides an ideal setting for examining how policies that shift the riskiness of choices impact gender gaps in real world outcomes.

Using a difference-in-differences (DiD) framework, we compare the behavior of men and women before and after the policy and provide three sets of results.

First, the policy induced a substantial increase in the gender gap in regrade requests. Although both men and women became more likely to request a regrade after the policy change, men responded much more strongly to the reform. The gender gap in regrade requests increased by 90%. By examining students who originally received F grades, which cannot be marked down, we show that the increased likelihood of a negative regrade outcome contributes to the gender gap.²

Second, the gender gap in requests translates into a gender gap in real world downstream outcomes: grades earned via regrade. This implies that the ostensibly gender-neutral regrade policy may push qualifying women below entry thresholds imposed by employers. For example, prior studies have found that a one-point increase in college grade point average (GPA) raises the employer's hiring interest by 0.8 standard deviations (SDs) (Kessler et al. 2019), that callback rates increase by seven percentage points when college GPA increases from 2.5 to 3.6 (Quadlin 2018), and that earnings increase with as much as 20% for students who are randomized into getting a 1-SD higher GPA (Landaud et al. 2024). This highlights the role that social institutions may play in shaping the dynamics of gender gaps in society.

Finally, why might men and women differ in their response to the change in probabilities of regrade outcomes? Two channels related to how people respond to risk are the most likely mechanisms for our effect: (1) risk preference-based channels, which deal with how men and women value different outcomes, and (2) belief-based channels, which deal with how likely men and women think different outcomes are. Prior literature has shown that men are less risk averse than women (Croson and Gneezy 2009, von Gaudecker et al. 2011, Charness and Gneezy 2012, Niederle 2015, Filipin and Crosetto 2016, Buser et al. 2021, Thöni and Volk 2021). They are also less sensitive to losses or bad outcomes than women (von Gaudecker et al. 2011). The literature has also shown that men are more confident than women and have more optimistic beliefs about their own performance and outcomes as a result (Barber and Odean 2001, Niederle and Vesterlund 2007, Coffman et al. 2021, Exley and Kessler 2022, Exley and Nielsen 2024). Theoretically, both the preference-based and the belief-based channels could be relevant in our context because they deal explicitly with probabilities. We use several methods to examine the relative contribution of these mechanisms to our core findings, including structural modeling, heterogeneity analyses by original grade, belief updating, and online experiments. In our experiment, we are able to directly elicit risk preferences, confidence, and beliefs. Consistent with the field data, beliefs and confidence emerge as stronger predictors of the demand for regrades in our experimental data. Taken together, our mechanism analyses show that gender differences in preference-

based risk mechanisms such as risk aversion and loss aversion are not sufficient to generate the effects we identify and that our data are more consistent with belief- and confidence-based mechanisms.

The key contribution of this paper is to show how ostensibly neutral policies can generate gaps across groups who differ in their risk-taking behavior. We use unique individual-level student data and exploit a policy that changed the riskiness of educational choices holding all other elements constant. This allows us to provide novel evidence on the consequences of gender differences in the response to a change in the likelihood of outcomes in a real-world decision environment. Our mechanism analyses show that gender differences in beliefs and confidence likely drive the results. We contribute to the existing literature in several ways.

First, we build on the well-established experimental literature that examines gender differences in preferences and beliefs. A key focus of this literature has been to examine how laboratory measures of preferences and beliefs correlate with real world behaviors and judgments (Sutter et al. 2013, Falk et al. 2018). Examples include preferences related to risk preferences (Dohmen et al. 2011), time preferences (Meier and Sprenger 2010), and competitiveness (Buser et al. 2014, Sutter and Glätzle-Rützler 2015). Beyond overall correlations, there also has been a growing interest in understanding the way that demographic differences in preferences lead to differences in outcomes, particularly focused on the role of gender (Baldiga 2014, Buser et al. 2014, Coffman and Klinowski 2020, Murciano-Goroff 2022, Cortés et al. 2023, Li and Zafar 2023, Coffman et al. 2024). Because of data limitations and empirical challenges, much less work has examined the causal role that specific differences in risk preferences and beliefs play in driving behavior outside the laboratory (Flory et al. 2015, Samek 2019). In this paper, we isolate the role of riskiness in a real-world decision environment, thereby advancing our understanding of how systematic differences in risk preferences and beliefs contribute to the demographic gaps that we observe in the world. The precise impact of the policy on the probabilities of the outcomes allows us to make strong causal claims about the impact of increased riskiness. A variety of mechanism tests point to gender differences in beliefs as the primary driver of the gender differences we document.

Second, there is a rich empirical literature documenting gaps in key labor market outcomes across groups (gender, ethnicity, age) and trying to identify channels through which these gaps occur (Altonji and Blank 1999, Goldin and Rouse 2000, Bertrand 2011, Olivetti and Petrongolo 2016, Blau and Kahn 2017). These studies have well-identified reduced-form effects and have greatly contributed to our understanding of existing gaps in key labor market outcomes across groups.

However, data limitations have often prevented these studies from relating these gaps to differences in specific traits or behaviors. This is understandable, as such exercises would require two distinct sources of variation (one that allows identification of the reduced-form effects and one that identifies how a particular character trait interacts with that policy). We build on this literature by evaluating a policy that exclusively shifts one component of a decision: the riskiness of making a choice. This allows us to advance the empirical literature by shedding light on whether specific cross-group behavioral differences contribute to the education and labor market gaps we observe in the real world.

Finally, this paper expands on an impressive and more specific literature that focuses on gender differences in labor market outcomes. In this literature, there exists a rich set of studies documenting persistent and prevalent gender gaps in labor market outcomes that cannot be explained through traditional mechanisms such as demographics (education, skills, ability), discrimination, or preferences for specific jobs (Blau and Kahn 2017). As a consequence, a large set of papers have focused on trying to identify differences in behaviors and preferences between men and women to see if this can account for some of the remaining gaps. This includes differences in competitiveness (Niederle and Vesterlund 2007, Sutter and Glätzle-Rützler 2015, Tungodden and Willén 2023), self-promotion and confidence (Coffman et al. 2021, 2024; Exley and Kessler 2022; Murciano-Goroff 2022; Exley and Nielsen 2024), bargaining ability (Card et al. 2013), asking (Babcock and Laschevar 2003, Small et al. 2007, Leibbrandt and List 2015, Coffman and Klinowski 2020, Li and Zafar 2023, Roussille 2024), and role models (Riise et al. 2022). The majority of these studies have been conducted in the laboratory.

Our contribution to this literature is to document a substantial difference in the behavioral response of men and women to a real-world change in the riskiness of choices through an application to the education context. Although our analysis is focused on the education context, we believe that this finding generates a broader set of takeaways for policymakers as well as researchers focusing on gender differences in labor market outcomes. Specifically, our findings illustrate that government and firm policies can have unintentional effects on gender gaps in outcomes by altering the decision environment by inducing men and women to take up opportunities differentially due to pre-existing gender differences in risk preferences and beliefs or confidence. The policy we study did not target men and women differentially, but because men and women differ in their beliefs and risk preferences, they responded to the policy in different ways. In turn, they faced different downstream implications in terms of grade point changes. On a broader level, this

highlights that some of the gender gaps we observe in the labor market may be driven by unintended consequences of government policy and management decisions through its interaction with gender differences in risk preferences and beliefs or confidence. We believe this is a key finding that should be emphasized in the policy arena in the future, both on a national level as well as on a firm level.

We use the regrade environment at the Norwegian School of Economics (NHH) to isolate, and provide novel evidence on, the effect of changing the riskiness of choices (i.e., altering the probabilities of the outcomes associated with choices) on the gender gap in real world outcomes. One other fascinating study has looked at the regrade behavior of men and women at universities: (Li and Zafar 2023). However, although the setting in itself is interesting (regrade behavior at universities), the main contribution of our paper is our ability to isolate the effect of changes in the riskiness of choices on gender differences in real world behavior. This is a question that Li and Zafar (2023) are unable to address due to data constraints. Specifically, they do not have an exogenous source of variation in the riskiness of choices, and they cannot observe the regrade requests that students made, only grade adjustments entered in the university system.

Regarding generalizability to other settings and contexts, it is important to note that the goal of our paper is to isolate the effect of changes in the riskiness of choices on gender differences in behavior (decision to request regrades). That this gender difference in behavioral response translates into differences in downstream outcomes in the education setting is both noteworthy and of independent interest. However, the exact magnitude of this second-order effect will depend on a number of factors, including the size of the risk change and the particular setting that the individuals are in. As such, we believe that the core contribution of our paper is to demonstrate a gender difference in behavioral response to changing the riskiness of choices, show that this affects downstream outcomes in one particular setting, and encourage future research to explore the implication of this across different settings and contexts.

In terms of policy implications, our paper highlights the role of systematic differences in risk taking in driving gender disparities in education and labor market outcomes and highlights the role that changes in the riskiness of choices can play in generating disparities across subgroups. Our mechanism evidence suggests that beliefs and the possibility of negative outcomes are particularly important in driving gender gaps in risk taking. This is important not only for governments when designing and implementing large-scale public policies (that often generate shifts in the risk associated with individual choices but are designed to be ostensibly neutral across groups), but also for managers

when implementing the design of incentive and bonus schemes. Specifically, any such scheme that provides differential rewards for risky behaviors will likely benefit men over women, thus contributing to overall inequality among men and women even conditional on ability and skills. This suggests that the extent to which social institutions and work incentive schemes are designed to balance the risk and reward of individual choices may have substantial effects on how cross-group gaps are shaped in society.

The rest of the paper proceeds as follows. Section 2 discusses the study design and institutional background. Section 3 discusses our data and methods. Section 4 presents the results from the administrative data. Section 5 presents the experimental results. Section 6 discusses our findings and concludes.

2. Study Design

2.1. Institutional Background

Our study is conducted using data from the NHH. NHH is a public business school with free undergraduate and masters programs in business. The undergraduate program is ranked among the most popular programs in the country, with an acceptance rate of around 9%. The school has approximately 3,500 students and attracts students from all parts of Norway as well as internationally.

Although the main contribution of our paper is to isolate the effect of changes in the riskiness of choices on gender differences in behavior, examining this question in an educational setting with respect to grades is particularly interesting. First, grades are instrumentally valuable for future career motives and serve an important role in securing scholarships, internships, and entry-level jobs (Quadlin 2018, Kessler et al. 2019, Landaud et al. 2024). For example, employers in Norway often set GPA thresholds that have to be passed in order to qualify for their internship programs (Landaud et al. 2024). Second, grades can enter directly into the utility function as value carriers or signals about own ability. This second motive may not be independent of their instrumental value but may act as an additional motive for students to want to improve their grades. For example, students tend to care about their grades even if they have limited instrumental value (e.g., certain U.S. business schools have grade non-disclosure, yet students still exert substantial effort in their courses; Floyd et al. 2024).

We recognize that generalizability is always difficult to speak to when examining evidence from relatively specific settings. The focus of our paper is on how changes to the decision environment can affect gender gaps through gender differences in response to risk. Although we use a Norwegian educational setting because it has useful identifying features, our work builds on a broad literature on gender differences on a variety of dimensions relevant to risk, specifically

gender differences in risk preferences and gender differences in confidence. In addition, prior work conducted at NHH has shown that gender differences in competitiveness and other behavioral traits are very similar to that in the broader student population (Almås et al. 2016, Buser et al. 2021, Tungodden and Willén 2023), and not much different from those in the broader nonstudent population either (Schroyen and Aarbu 2018). We think the results are not about the regrade decision specifically but about how men and women respond differently to the opportunity to take up a risky decision when that risk changes. Additionally, we observe similar gender gaps in both our Norwegian sample and an online sample of American workers, suggesting that our results are not confined to the elite university setting and may be broadly relevant for policy makers and firms altering the risk of various decision environments.

2.2. Grading and the Regrade Process

Grading at NHH is done on a six-point scale that ranges from A through F, with F representing a fail and A representing the top grade. There is no cap on the percent of students that can receive a specific grade. Certain courses use a binary “pass-fail” scale, but this is rare.

All students have the right to request a regrade of any assignment including exams, with the exception of oral exams. The appeal process begins with the student requesting an explanation for the grade received. This request must take place within seven days of having received the original grade and is done either through the online student learning interface or via email to the exam office. The original examiner has two weeks to provide an explanation for why a certain grade was given. The examiner may choose whether to give the student a written or an oral explanation. There are no additional costs to submitting a justification requests.

After receiving the original examiner’s explanation, the student can request a regrade of the assignment by another anonymous examiner. In our analysis, we focus on the decision to request a regrade through this step.³ The deadline for submitting an appeal is three weeks after receiving the explanation. The exam is sent to new examiner(s), who were not involved in issuing the original grade assigned to the student. There is no cost to requesting a regrade beyond submitting a request. Importantly, both the original examiner and the new examiner are blind to the gender of the student at the time of grading. The result of the appeal is final. Prior to the 2014 reform, the new examiners received the student’s initial grade, the original examiner’s justification for awarding the student that grade, and the student’s motivation for requesting a justification.

Following the national legislative change, which was passed in June 2014 and impacted all higher education

institutions in the country, the new examiners would not receive any of this information. The new examiners were thus completely blinded from the prior assessment decision. The goal of the national regrading reform was to provide students with a fair chance of reassessment that was not anchored by prior results. By blinding the new assessor from the student’s original grade, policy-makers believed they could provide a more transparent process in which the new assessor did not simply rely on the original grade when reassessing the assignment, but rather provided a grade that reflected the new assessors objective belief about the quality of the assignment. Although there was a considerable public debate on the question of whether to blind the regrade process or not, the National Student Organization of Norway was a strong proponent of the policy.⁴

3. Data and Method

3.1. Administrative Data

We exploit unique administrative data from NHH. This data extends from 2009 through 2019 and consists of information collected and combined from five separate student registers: the student population register, the student major register, the student course register, the student grade register, and the student regrade register.

These data provide us with individual-level student information on a set of demographic characteristics (gender and age). It also provides us with detailed information on the students’ field of study, which student cohort they belong to, and which courses they have taken at the school.

Crucial to our analysis is the ability to observe the students’ performance on every exam and assignment in every course during their time at the school, whether the student requested a regrade, and what the final outcome of the regrade was. We collect this information from the student grade and regrade registers. For our main analyses, we restrict the data to all grades in the range B to E because only these grades can experience both a grade increase and a grade decrease.⁵ However, in Section 4.1, we use the F grades as a mechanism test for downside risk. One benefit of the Norwegian setting is the formalization of the grade and regrade process, ensuring that we observe all regrade requests made during a semester including those for which the final grade is unchanged, thus removing any concerns of selective reporting.

Our final sample includes all assignments from 12,684 unique students. Forty percent of the students are women. The average year of birth is 1990. A total of 8,979 students are observed only in one program (i.e., bachelors or masters), and 3,705 are observed in both programs. Average grades are 3.52 on a scale from 0 to 6; 45.1% of students in our sample ask for a regrade. Summary statistics of our administrative analysis

sample are provided in Panel A of Table A1 in the Online Appendix. Additionally, Figure A1 in the Online Appendix shows that the cumulative distribution functions of yearly average grades among men and women. Overall, men and women have relatively comparable grades, and even though men have slightly better grades than women at the upper end of the distribution, the differences are small.

3.2. Estimation Strategy

To examine if the policy induced a change in the gender gap in regrades, we rely on the following difference-in-differences specification:

$$Y_{ict} = \alpha_0 + \beta_1 Female_i + \beta_2 Post_t + \beta_3 (Female_i * Post_t) + \rho_t + \gamma_c + \tau_g + \epsilon_{ict}, \quad (1)$$

where Y_{ict} is one of our outcomes (such as, for example, a binary outcome for whether a regrade was requested) for student i in class c and year t . A student may have multiple assignments per course. *Female* is a binary variable taking the value of one if the student's sex is female. *Post* is an indicator variable for if the observation is from the postreform period (≥ 2014). The coefficient β_1 provides information on the difference in outcomes between men and women prior to the reform, β_2 provides information on the effect of the policy on male regrade behavior, and β_3 identifies whether the policy had a differential effect on female students. The baseline specification includes no fixed effects. Additional specifications include a full set of year (ρ_t), class (γ_c), and original course grade (τ_g) fixed effects. Our findings are not impacted by the inclusion of these fixed effects, which is something we show in Section 4.

The assumption underlying our estimation strategy is that men and women were not trending differently in their outcomes prior to the policy, such that the pre-reform gender gap can be used as a plausible counterfactual of what the gap would have looked like in the postreform period absent the policy. In addition, there can be no other shocks contemporaneous with the regrade reform that is correlated with the outcomes and that differentially impact men and women.

To examine the plausibility of the common trends assumption, we estimate a set of event studies:

$$Y_{ict} = \alpha + \sum_{t=2009}^{t=2019} [\delta_t(Female_i)] + \rho_t + \gamma_c + \tau_g + \epsilon_{ict}, \quad (2)$$

where δ_t traces out any pretreatment relative trends across gender ($t = 2009$ through $t = 2013$) as well as any time-varying treatment effects ($t = 2014$ through $t = 2019$). We omit 2013, so all effects are relative to that year. Examining potential time-varying treatment effects are particularly interesting as students may adjust their expectations over time following the implementation of the policy. This is something we examine directly in the

next section. Note that the policy was implemented halfway through 2014. We therefore expect any potential effect in 2014 to be smaller since only half of the year was subject to the treatment.

The second assumption, that there are no other shocks contemporaneous with the regrade reform that is correlated with the outcomes and that differentially impact men and women, is more difficult to test. However, we note that the school did not implement any other large-scale policies at the time of the regrade reform. In addition, we present results from a series of robustness and sensitivity analyses in which we alter the sample, the model specification, and the fixed effects. Further, we replicate our main findings from the administrative data analysis in an online experiment in which we can eliminate any risk of other contemporaneous shocks driving the results. The robustness of our results across these specifications, and the generalization of our findings to an online experiment, strengthen the causal interpretation of our findings.

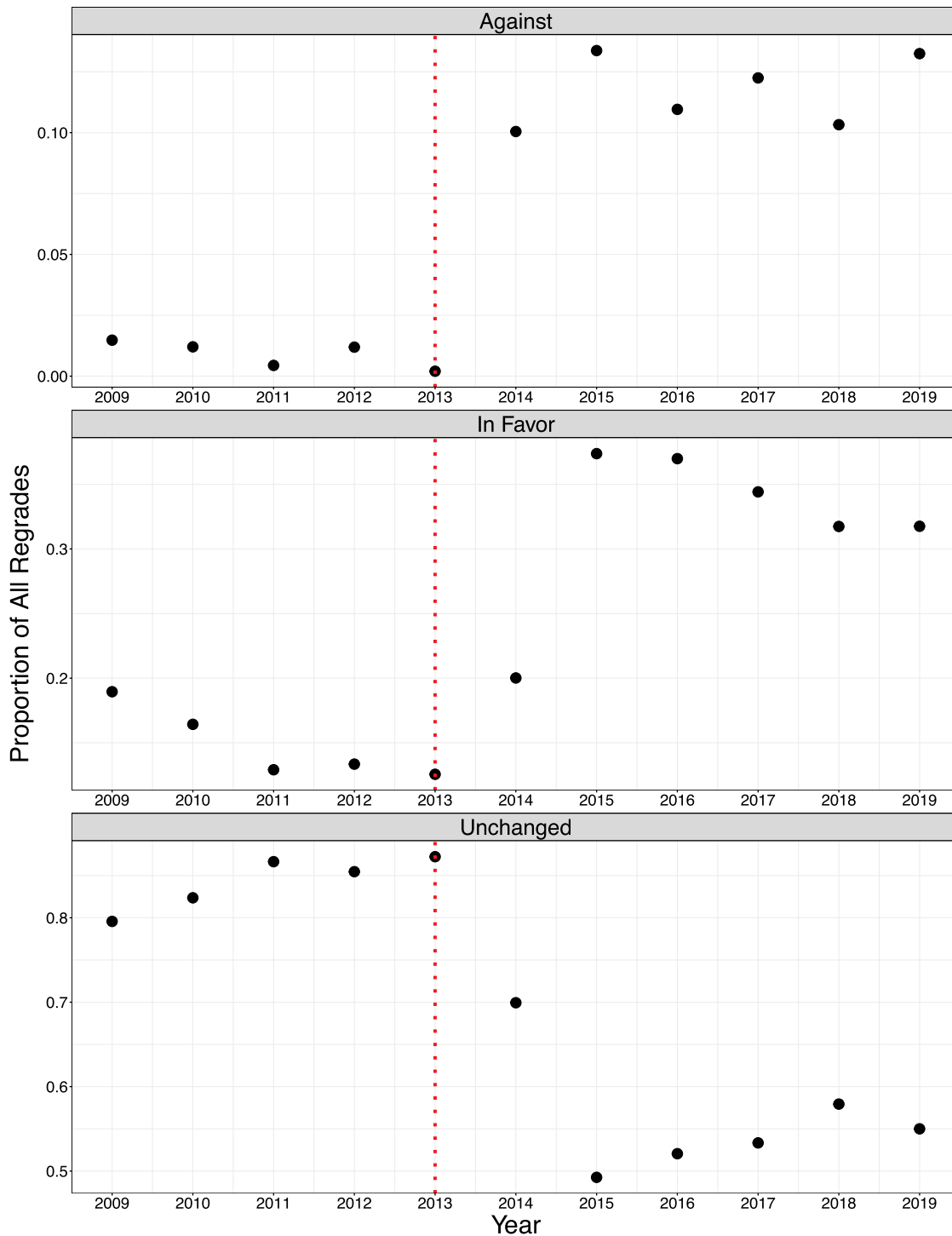
4. Administrative Data Results

4.1. Main Findings

In this section, we present our key findings on the effect of risk on the gender gap in real world choices. We establish that the expected value and the risk of requesting a regrade changed substantially following the new regrade policy. We then examine how this policy affected the probability of regrade requests across men and women, and if this translates into gender differences in grade points. Finally, we turn to mechanisms. Theoretically, both preference-based and the belief-based channels could be relevant in our context, since they deal explicitly with probabilities. We use several methods to examine the relative contribution of these mechanisms to our core findings, including structural modeling, heterogeneity analyses by original grade, belief-updating, and online experiments.

4.1.1. Regrade Risk. Figure 1 shows the final outcome of all regrade requests over time, from 2009 through 2019. Figure 1(a) shows the share of regrade requests that results in a worse grade than the initial grade, Figure 1(b) shows the share of regrade requests that results in a better grade than the initial grade, and Figure 1(c) shows the share of regrade requests that result in no grade change.

The share of regrade requests that were marked up, marked down, or not changed was stable in the pre-policy years, and there are no noticeable time trends in the final outcome of regrade requests before 2014. This suggest that there were no contemporaneous policies in the grade and regrade process in the years leading up to the reform that altered the incentives to request regrades, which could contaminate our findings.

Figure 1. (Color online) Proportion of Regrade Outcomes by Year for All Requests

Notes. Proportion of grades submitted for regrading that are decreased, increased, and left unchanged over the period 2009–2019. The dotted line denotes 2013, the last full year before the new regrade policy went into effect.

Prior to the policy change, 84.4% of all regrade requests resulted in an unchanged grade, 14.7% resulted in a grade improvement, and only 0.9% led to a grade reduction. The expected return on asking for a regrade was 0.14 grade points. In other words, the vast majority

of regrade requests prior to the reform resulted in no change of the initial grade.

After the reform, the share of regrades that receive a negative reevaluation increases to 11.7% of all requests and the share of regrades that receive a favorable

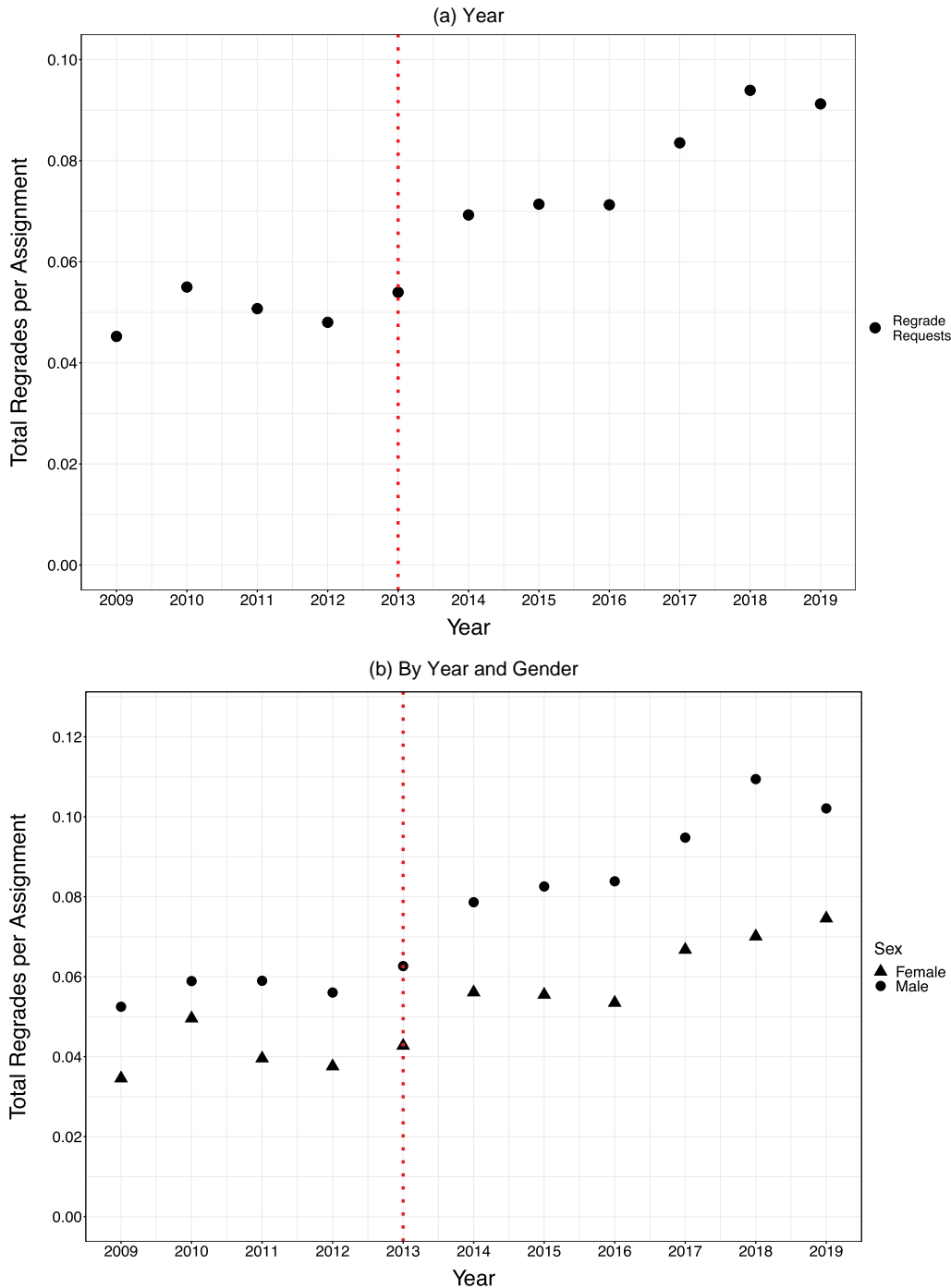
reevaluation increases to 32.2%. The expected return on asking for a regrade in the postreform period was 0.22 grade points, a 50% increase.

Taken together, Figure 1 demonstrates that the regrade policy generated a shift in the expected value as well as the probabilities associated with different regrade outcomes. We use the change in these objective probabilities

to examine whether altering the riskiness associated with a regrade request has an impact on the choices made by students, and if that impact differ across men and women.

4.1.2. Regrade Requests. In Figure 2, we present evidence on the extent to which the change in probabilities

Figure 2. (Color online) Proportion of All Grades Submitted for a Regrade



Notes. Proportion of all grades submitted for regrade requests plotted by year in (a) and by year and gender in (b). The dotted line denotes 2013, the last full year before the new regrade policy went into effect. In (b), the filled in dots show the proportion of assignments by male students submitted for regrades and the hollow diamonds show the proportion of assignments by female students submitted for regrades.

of grade changes following regrade events altered the regrade request probability, both overall (Figure 2(a)) and by gender (Figure 2(b)). Figure 2(a) demonstrates that the likelihood of requesting a regrade for an assignment was relatively stable at around 5% of all assignments prior to the reform. After the reform, this likelihood increased to around 8%: a 60% increase relative to baseline. Students responded strongly to the reform, suggesting that the change in the expected value of a regrade request substantially increased the attractiveness of regrade requests.

Figure 2(b) provides three important results. First, prior to the reform there was a substantial gender gap in the propensity to request a regrade. Second, both male and female students responded to the reform by being more likely to request a regrade of an assignment. Third, the response of male students to the policy was much greater than the response of female students, such that the policy substantially augmented the pre-existing gender gap in regrade requests.

To assess the gender gap in regrade requests, Figure 3 plots point estimates of the interaction between female and time obtained from estimating Equation (3). Figure 3(a) provides results for all assignments and Figure 3(b) provides results focusing exclusively on written exams. Both panels illustrate that the introduction of the regrade policy generated a statistically significant and economically meaningful increase in the gender gap in regrade requests, with male students becoming much more likely to request a regrade relative to female students after the policy change.

To summarize the results from the event studies and examine robustness across different specifications, Table 1 shows the results from estimating Equation (2) on the probability of asking for a regrade. In Panel A, we show results for all assignments. In Panel B, we provide results obtained only based on the written exams. In each of the panels, we show how the results change as we sequentially add layers of fixed effects, from none (column (1)) to a full set of year, class, original grade, and whether the course was retaken fixed effects (column (4)).⁶

The results demonstrate that there was a significant gender difference in regrade requests even before the policy was implemented of approximately 1.7 percentage points (coefficient on *Female*). The results further show that the reform induced a significant increase in the probability of requesting a regrade, up 3.4 percentage points in the baseline specification (coefficient on *Post Blinding*). Finally, the results demonstrate that this increase was significantly lower for female students by 1.2 percentage points (coefficient on *Female × Post Blinding*). Taken together, the results demonstrate that changing the probabilities of regrade outcomes (i.e., altering the risk of requesting a regrade) generated a 90% increase in the gender difference in regrade requests.

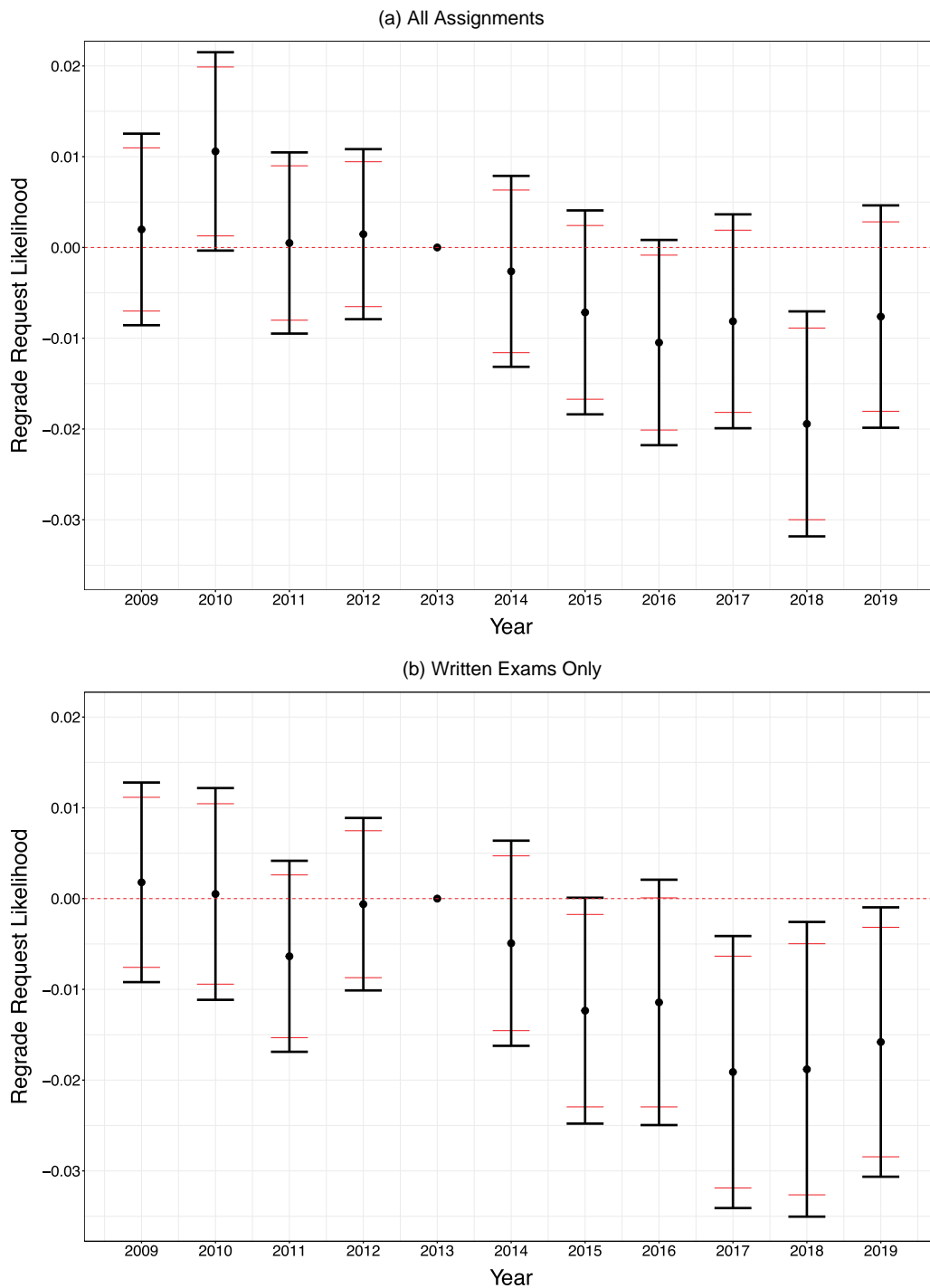
To understand the extent to which gender differences in response to grade decreases are generating the gender gap we find, we examine F grades. F grades cannot be decreased, regardless of the regrading regime. The average return for requesting a regrade on an F increases from 0.22 grade points before the policy change to 0.46 grade points after the policy change; an increase in expected value with no possible downside risk. If gender differences in response to downside risk are an important contributor to our effect, we would expect no gender differences for F grades because there is no possibility of downside risk. We present the results by original grade in Table 2. We find no gender difference in the likelihood of requesting a regrade conditional on initially receiving an F in the preperiod, nor do we find a significant difference-in-differences estimate (though there is a strong overall increase in the likelihood of regrade requests after the policy change).

To better understand which students are impacted by the policy, we consider how the 2014 policy change affects gender gaps both on the extensive and intensive margins using two different analyses.⁷ First, we observe 8,983 students in just the pre- or postperiod and 2,809 students in both periods in our main analytic sample. For the extensive margin, we classify each student as having submitted a regrade if we ever observe them doing so (regardless of whether this was done in the preperiod or in the post-period). This means that a student we observe both in the pre-period and in the post-period who requests their first regrade in the post-period is still classified as having asked in our observation of them in the pre-period. We view this classification as conservative because these students, who might have been impacted by the policy change, are not classified as such in the extensive margin analysis. For the intensive margin, we examine the number of grade requests among the people classified as having ever asked split by pre and post grade change, capturing the effect of the policy on the intensive margin.

Female students are seven percentage points less likely than male students to ask for a regrade in the pre-reform period. The gender gap on the extensive margin widens in the postreform period by 3.5 percentage points, or about a 50% increase in the gender gap in ever asking for a regrade. Turning to the question of the intensive margin, we condition on students who ever ask for a regrade. There is an initial gender gap of 0.16 requests which increases by 0.18 requests (110%) in the post reform period. Even among female students who are willing to ask at all, they ask for fewer regrades than male students do. These effects are not explained by differential changes in the average grade submitted for a regrade after the policy, although the average grade does decline (Figure A3 in the Online Appendix).

In the Online Appendix, we examine whether there are heterogeneous effects in the gender gap based on

Figure 3. (Color online) Event Study Plot for the Interaction Between Year and Female



Notes. Interaction coefficients δ_t for all exams in (a) and written exams in (b). The dotted line denotes the zero point, which can be interpreted as the gender gap in 2013, the last full year before the policy was implemented. The wider error bars denote 95% confidence intervals and the narrower error bars denote 90% confidence intervals. Standard errors are clustered at the student level. The policy was implemented in the middle of 2014 so we expect a smaller effect in that year.

the stakes of the exams and prior grades. First, it is theoretically plausible that the regrading behavior is more pronounced when the stakes are high. However, if we stratify the sample based on whether the course grade was determined by a single assignment (or exam) or by

multiple assignments (or exams), we do not see a difference in the gender gaps (Table A5 in the Online Appendix). Second, it might be the case that the policy response of students depend on their prior grades and the extent to which a specific assignment score deviates

Table 1. Effect of Policy on Regrade Request Probability Using Simple DiD Estimates

Variable	Model 1	Model 2	Model 3	Model 4
Panel A: DiD estimates for all assignment types				
<i>Intercept</i>	0.058*** 0.001			
<i>Female</i>	−0.017*** 0.002	−0.017*** 0.002	−0.016*** 0.002	−0.016*** 0.002
<i>Post Blinding</i>	0.034*** 0.002			
<i>Female × Post</i>	−0.012*** 0.003	−0.012*** 0.003	−0.011*** 0.003	−0.011*** 0.003
Observations	199,253	199,253	199,253	196,946
Fixed effects: Year		X	X	X
Fixed effects: Class			X	X
Fixed effects: Original grade			X	X
Fixed effects: Retake				X
Panel B: DiD estimates for written exams				
<i>Intercept</i>	0.047*** 0.002			
<i>Female</i>	−0.015*** 0.002	−0.015*** 0.002	−0.014*** 0.002	−0.014*** 0.002
<i>Post Blinding</i>	0.042*** 0.003			
<i>Female × Post</i>	−0.013*** 0.003	−0.012*** 0.003	−0.012** 0.004	−0.011** 0.004
Observations	132,733	132,733	132,733	131,027
Fixed effects: Year		X	X	X
Fixed effects: Class			X	X
Fixed effects: Original grade			X	X
Fixed effects: Retake				X

Notes. Simple difference in difference estimates of the likelihood of submitting a regrade request. Panel A presents the estimates for all assignment types. Panel B presents the estimates for just written exams. Column one shows the model without fixed effects, column two includes year fixed effects, column three includes year, class identifier, and original grade fixed effects, and column four includes year, class identifier, original grade, and a fixed effect for whether the exam was the student's second time taking an exam in that class code. Standard errors are presented below the estimates and are clustered at the student level.

⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

to the average grade of the student. To this end, we look at heterogeneity by students' prior grades in two supplemental analyses. We control for (i) students' prior average grade and (ii) the deviation of the current grade from the students' prior average grade. Next, we

split our sample into two groups based on whether the student had above or below median average grades in the first year we observe them in the data. The results from these exercises are provided in Tables A6 and A7 in the Online Appendix. Controlling for prior grades

Table 2. Effect of Policy on Regrade Request Probability Using Simple DiD Estimates for Each Grade Level

	F grades	E grades	D grades	C grades	B grades
<i>Intercept</i>	0.064*** 0.008	0.023*** 0.004	0.091*** 0.004	0.083*** 0.003	0.032*** 0.002
<i>Female</i>	−0.007 0.012	0.001 0.005	−0.009 0.006	−0.023*** 0.004	−0.016*** 0.002
<i>Post Blinding</i>	0.076*** 0.014	0.063*** 0.007	0.103*** 0.007	0.058*** 0.004	0.001 0.002
<i>Female × Post</i>	0.020 0.024	0.002 0.011	−0.017 ⁺ 0.010	−0.023*** 0.006	−0.003 0.002
Observations	4,473	8,554	24,764	70,711	95,224

Notes. Simple difference in difference estimates of the likelihood of submitting a regrade request for each grade level without fixed effects. Standard errors are presented below the estimates and are clustered at the student level. Event study estimates are provided in Figure A9 in the Online Appendix.

⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

do not substantially affect the size of the interaction coefficient between treatment and gender, and both above and below median students' regrade requests respond similarly to the policy change. These results imply that the gender gap in regrade requests that we observe is a more general phenomena that extends to men and women across the entire GPA distribution.⁸

4.1.3. Downstream Outcomes. Next, we ask whether the augmented gender gap in regrade requests has downstream implications for the students, both in terms of average yearly grade changes from regrades and the average grade point change per regrade request. We present results on these questions in Figure 4.

The reform is associated with a significant increase in the gender gap in average yearly grade change, going from no gender gap in 2013 to a gender gap of approximately 25% in 2016 (Figure 4(a)). Figure 4(b) shows the average grade point change earned by male and female students. If anything, this panel shows female students earn slightly more grade points per request, but it does not appear that male or female students are significantly better at identifying grades to submit for reassessment. For simple difference-in-differences estimates, see Table 3. Consistent with the increase in the change in points earned via regrade but no change in returns conditional on request, grade increases are more likely for male students after the policy but so are grade decreases.⁹ We also find that total regrade requests is correlated with graduate study probability (correlation coefficient of 0.016, with a standard error of 0.003), such that the identified gender difference in behavior may have downstream implications beyond the direct grade points effect. However, this is suggestive evidence with weaker causal interpretation than our finding related to grade points.

To better understand how disparities in grade requests in our setting contribute to a gender gap in academic achievement, we also have estimated our main model using the difference between average yearly grades computed with the original grade received and the final grade after regrade in the course as an outcome (Table A8 in the Online Appendix). Overall, the results from this analysis demonstrate that the regrading reform increased the gender gap in the difference between the original and the final grade in the course by 100%. Although the absolute magnitude of this effect is relatively modest, it is large as a percentage of the pre-policy gap. We believe that this is a helpful alternative way of demonstrating the impact of the reform on the gender gap in academic achievement. Interestingly, these results also are in line with but larger than the impacts on GPA reported in Li and Zafar (2023); the only other paper that has examined gender gaps in regrade requests. Additionally, because men have slightly better grades on average,

this implies that the policy augments, rather than narrows, gender differences in average grades.

4.2. Mechanisms

The policy we use to isolate the effect of changes in the riskiness of choices on the gender gap in real world outcomes only impacts the probabilities associated with regrade outcomes while holding the rest of the decision environment constant. As such, the only mechanisms that can plausibly explain our findings are (1) gender differences in risk preferences and loss aversion, and (2) gender differences in beliefs and confidence about the likelihoods of different regrade outcomes. Both these alternatives deal with how people respond to uncertainty. In this subsection, we carefully examine these mechanisms and whether they are sufficient to produce the effects we observe. In the next section, we consider and rule out alternative mechanisms and explanations that are theoretically plausible but highly improbable in our setting.

4.2.1. Gender Differences in Preferences (Risk Preferences and Loss Aversion).

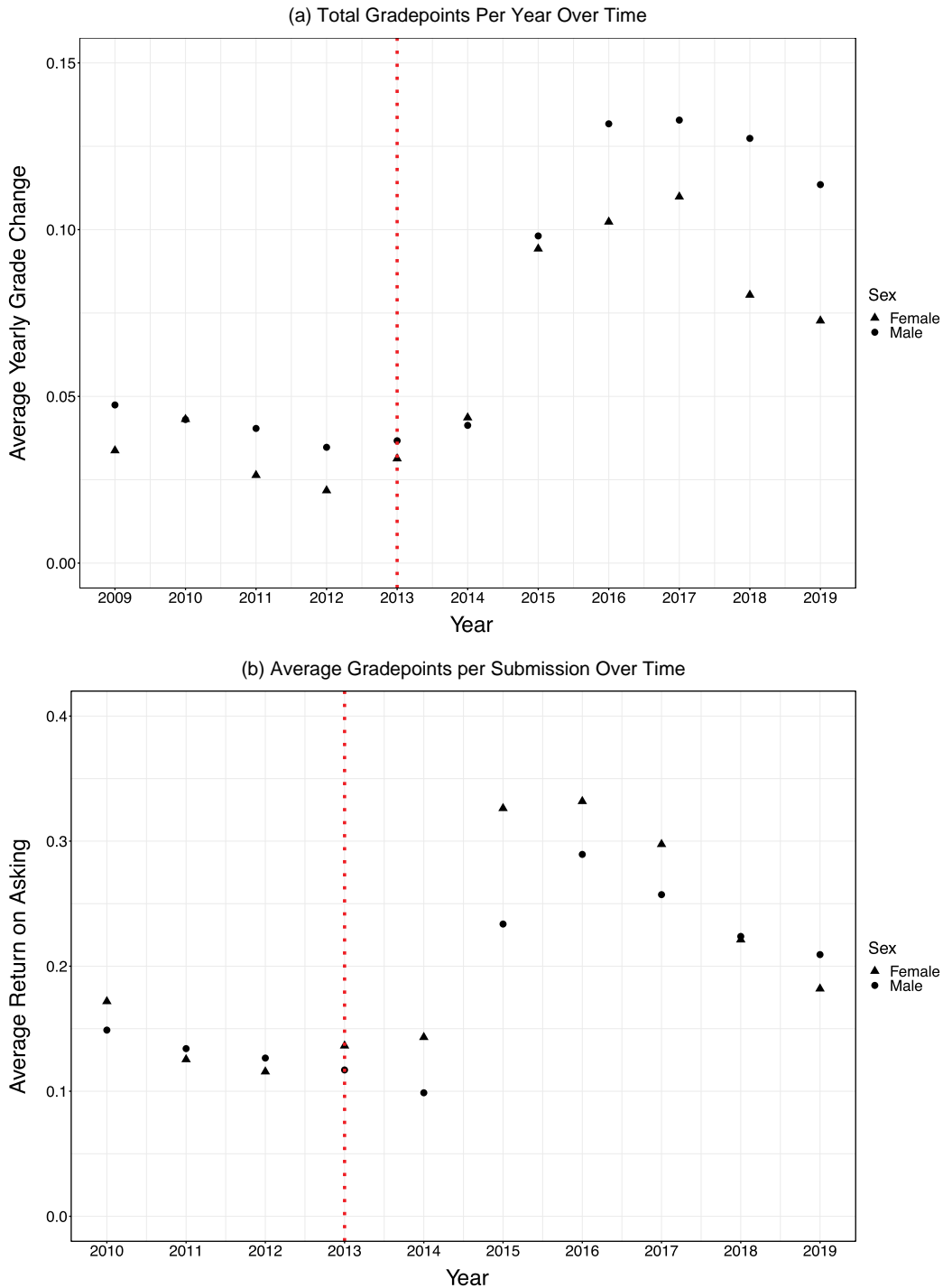
Prior literature on gender differences in risk preferences has shown that women tend to be more risk averse than men (Croson and Gneezy 2009, von Gaudecker et al. 2011, Charness and Gneezy 2012, Niederle 2015, Filippin and Crosetto 2016, Buser et al. 2021, Thöni and Volk 2021). In our setting, such gender differences could potentially explain why men respond more strongly to the regrade reform relative to women.

To examine this potential mechanism in detail, and its ability to explain our core findings, we consider two forms of gender differences in risk taking: (1) gender differences in risk preferences and (2) gender differences in loss aversion (that men weight losses less in utility terms than women).

4.2.1.1. Risk Preferences. If women are more risk averse than men, then they may ask for regrades relatively less than men after the policy change. This is because the policy increases the variance of outcomes, even though it also increases the expected value. As such, differences in risk preferences between men and women provides a potential explanation for our findings.

To examine whether this mechanism on its own can explain the increase in the gender gap in response to the policy, we estimate a simple structural model using nonlinear least squares:

$$\begin{aligned} \text{Min} \sum_j & (Choice - (EU(request) - EU(norequest)))^2 \\ & = \sum_j (Choice - ((\pi_{up}(x + 1)^\alpha) + (\pi_{down}(x - 1)^\alpha) \\ & + ((\pi_{nc} - 1)(x)^\alpha)))^2, \end{aligned} \quad (3)$$

Figure 4. (Color online) Average Grade Points per Year Earned Through Regrades and Average Return to Submitting a Regrade Request for Male and Female Students

Notes. Returns to asking for a regrade split by gender. (a) Average grade points changed in each year by male and female students over time. (b) Average return to a regrade submission in each year split by male and female students. In both panels, the circular dots represent male students, and the triangles represent female students.

where choice is a binary variable for whether a student asked for a regrade, x is the numeric current grade, π represents the probability of an increase, a decrease, or no change, and α captures the curvature of the utility

function (which represents a standard economic definition of risk preferences). Provided that people have correct beliefs about the aggregate likelihood that regrades have positive versus negative outcomes, we

Table 3. Change in Grade Points and Returns to Regrades

	Grade point change per year	Count of up changes	Count of down changes	Returns per request
<i>Intercept</i>	0.040*** 0.002	0.042*** 0.002	0.002*** 0.000	0.140*** 0.007
<i>Female</i>	-0.009** 0.003	-0.010** 0.003	-0.001 0.001	0.006 0.012
<i>Post Blinding</i>	0.067*** 0.005	0.113*** 0.005	0.055*** 0.002	0.081*** 0.011
<i>Female × Post</i>	-0.015* 0.007	-0.035*** 0.007	-0.022*** 0.003	0.021 0.020
Observations	39,443	39,443	39,443	13,882

Notes. DiD estimates on the returns to regrading. Column (1) presents the average change in grade points per year. Column (2) presents the count of changes that increase a student’s grade per year. Column (3) presents the count of changes that decrease a student’s per year. Column (4) shows the average return per regrade request conditional on asking for a regrade. Standard errors are presented below the estimates and are clustered at the student level.

⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

can interpret gender differences in risk preferences as being sufficient to explain the patterns we observe if the following two results are observed:

1. Women are more risk averse than men in both the pre- and postperiod.
2. Aggregate risk preferences are stable between the pre- and postperiod.

If we do not observe that women are more risk averse than men in both the pre- and postperiod, then there is no gender gap in risk preferences so those preferences cannot explain the differences in behavior. If we do not observe stable aggregate risk preferences, the instability is likely coming from model misspecification driven by other mechanisms that are impacted by the policy but not included in the estimation, such as gender differences in beliefs and confidence. We consider it unlikely that the policy change would directly impact students’ risk preferences since the population of students is very similar over time and the policy only impacts the probabilities associated with regrade outcomes.

Our results from this exercise are shown in Table A.10 in the Online Appendix. For men in the pre-period, we estimate that $\alpha = 0.69$, whereas for women, we estimate that $\alpha = 0.59$. That is, both men and women behave “as-if” they are risk averse in the pre-period, but men are significantly less risk averse than women. This suggests that reasonable estimates of risk aversion can rationalize the gender gap in the pre-period. In the post-period, we estimate that $\alpha = 1.08$ for men and that $\alpha = 1.03$ for women. Both these estimates imply risk seeking behavior and differ substantially from our preperiod estimates. The large changes in estimated risk preferences from the pre- to post-period are implausible, especially considering that other work at NHH has found standard gender gaps in risk preferences (Buser et al. 2021) and violate the second result we would require in order for gender differences in risk preferences to be sufficient to

explain our core results. Our conclusion from this simple structural exercise is, therefore, that our core DiD estimates are unlikely to be driven exclusively by gender differences in risk preferences. Because of the broad literature suggesting gender differences in risk preferences, ruling out this explanation is valuable even if it requires strong assumptions about beliefs.

4.2.1.2. Loss Aversion. Loss aversion (Kahneman and Tversky 1979) implies that people weight negative outcomes more than they do positive outcomes. The policy change we examine introduces the possibility of negative regrade outcomes, which were extremely rare in the pre period. As a result, if female students experience a greater degree of loss aversion (von Gaudecker et al. 2011), then the core results we present could potentially be explained by women placing increased weight on negative utility outcomes relative to men. To examine whether gender differences in loss aversion are driving the core results we found in Section 4, we investigate effect heterogeneity on regrade requests by original grade.

There are three particular heterogeneity analyses that can help shed light on the loss aversion mechanism in this setting. First, we examine gender differences in regrade requests among those who originally received an F grade. F grades cannot be marked down in the regrading process. If the increased gender gap is coming from loss aversion, we should therefore see no gender gap for Fs. Second, we examine gender differences in regrade requests among those who originally received an E grade. Going from an E grade to an F grade represents a considerably larger decline in utility than any other grade reduction because getting an F means that the student fails the course. This would require the student to either retake the exam, retake the course, or make up the credit hours in another course. Because loss aversion amplifies the impact of negative utility outcomes, if gender differences in loss aversion

were driving the gender differences in regrade requests that we find, we would expect to see the largest gender gap for Es.¹⁰ Finally, we examine gender differences in regrade requests among those who originally received a B grade. The share of grades that are reduced in the regrading process increases substantially for Bs relative to the other grades after the policy change, such that the expected value of asking for a regrade in the post period is actually negative for Bs, although it is positive for all other grades. If increased loss aversion was driving our results, we would expect to see a larger increase in the gender gap for B grades because of the high share of downgrades (Table A11 in the Online Appendix).

Table 2 presents the results for each original grade. First, in terms of F grades, the average return for requesting a regrade on an F increases from 0.22 grade points before the policy change to 0.46 grade points after the policy change; an increase in expected value with no possible downside risk. We find no gender difference in the likelihood of requesting a regrade conditional on initially receiving an F in the preperiod, nor do we find a significant difference-in-differences estimate (although there is a strong overall increase in the likelihood of regrade requests after the policy change). This supports the hypothesis that gender differences in the response to grade decreases are generating the gender gap we find and that that the observed gender disparity in regrade requests is not driven by men being more responsive to the increased potential for grade improvements than women.

Second, in terms of E grades, we also do not observe a statistically significant increase in the gender gap after the policy change. This is inconsistent with loss aversion being the main driver of our core findings. Specifically, if higher loss aversion for women was the main driver of the effect we find, we would have expected a large increase in the gender gap.

Finally, in terms of B grades, we see a small increase in the gender gap as a consequence of the policy, although it is not statistically significant. Because the share of Bs being downgraded increases substantially, loss aversion should be strong here. However, we do not see a large increase in the gender gap, a finding that speaks against loss aversion as the main driver of the results we identify.

Overall, our tests by original grade are not consistent with gender differences in loss aversion. Combined with our results on risk preferences from the structural estimation, it does not appear that either gender differences in risk aversion or gender differences in loss aversion are sufficient on their own for producing the difference in difference results we document.

4.2.2. Gender Differences in Confidence and Beliefs. As discussed above, the policy we examine only impacts the probabilities associated with regrade outcomes

while holding the rest of the decision environment constant, such that the only mechanisms that could plausibly explain our findings are (1) gender differences in risk preferences and loss aversion and (2) gender differences in confidence and beliefs about the likelihoods of different regrade outcomes. After having confirmed that neither gender differences in risk aversion nor gender differences in loss aversion are sufficient on their own to our results in the prior subsection, we now turn to examine confidence and beliefs.

In terms of beliefs and confidence, there is substantial evidence that men are more confident than women (Barber and Odean 2001, Niederle and Vesterlund 2007, Exley and Kessler 2022, Exley and Nielsen 2024). In our setting, this could help explain our core results through two different channels, both increasing men's beliefs about the expected value of asking for a regrade. First, men may just simply have more optimistic beliefs about expected values (a higher estimation of the likelihood of a positive regrade, a lower estimation of the likelihood of a negative regrade, or both) and they become more optimistic in the post-period. Second, men may have higher estimation of their own assignment quality. This would produce a similar effect on regrade likelihood beliefs and potentially would respond to the policy since the policy made grade changes more likely.

Although we do not have detailed data on beliefs, we can obtain suggestive evidence on the belief channel by examining gender differences in regrade behavior as a function of the original grade they received. These results are provided in Table 2 and show that men responded to the policy by requesting more regrades for all grades except Bs (for which we saw no change between pre and post reform even though the expected value of a regrade request of B grades decreases by 0.22 grade points). This pattern of results in our heterogeneity analysis is broadly consistent with increased optimism in the post-period among men.

Gender differences in confidence could be accentuated by the policy change because it changes probabilities of outcomes. It also introduces an additional margin, namely downside risk, that men might be more confident over. This would be consistent with our overall effect and with our results for every grade level except Es. For Fs, there is an increase in expected value, but no increase in downside risk. That is matched by an increase in the likelihood of regrade requests but no significant difference in the gender gaps post policy change. For D and C grades, we observe both an increase in the likelihood of requesting a regrade and the gender gap, consistent with both the increase in expected value and the increase in downside risk. For Bs, where expected value decreases, we do not see an increase in the likelihood of regrade requests. However, we do see a slight increase in the gender gap though it is not statically

significant. For Es, we observe an increase in the expected value after the policy change, and we also see an increase in the likelihood of requesting a regrade for both men and women. However, there is no significant gender gap, as discussed above.

One additional direct test for whether beliefs contribute to the regrade decision of students is to examine whether peoples' behavior responds to prior regrade experiences. That is, do people learn about the probabilities of outcomes from their own experiences and update their regrade request behavior? We cannot do this analysis in our DiD framework because there were almost no negative regrade outcomes in the pre-period. However, we can perform this analysis in the post-period and in the full data. Under the assumption that outcomes are as-if random draws from the outcome distribution, this analysis would identify the causal effect prior regrade outcome on the likelihood of asking for a regrade in the future.

In Table 4, we show the results on how a student's regrade requests respond to their first regrade outcome. Men increase their likelihood of regrade requests following a positive regrade outcome and decrease it following a negative regrade outcome, relative to an unchanged grade. These results are consistent with learning about the odds of a regrade. Women are less likely to ask for a regrade than men after an unchanged outcome, which is consistent with a gender difference in the confidence mechanism. Finally, a male student who gets a negative regrade outcome is 3.2 percentage points

less likely to ask for a regrade than one who receives an unchanged grade and a woman is 1.1 percentage points less likely. Although the interaction is not statistically significant, this difference is consistent with women having more pessimistic beliefs to start and thus being less surprised by the negative outcome.

Taken together, our results in Section 4.2 suggest that gender differences in risk preferences and loss aversion are not sufficient to explain the increased gender gap after the policy change. In addition, the results suggest that gender differences in confidence or beliefs are necessary to rationalize the data and more broadly consistent with the patterns we observe. We also directly observe evidence of learning from prior experience, which is consistent with an important belief-based channel for deciding whether to regrade and suggestive but weak evidence that women have more pessimistic priors.

In the next section, we discuss and rule out a variety of alternative mechanisms that researchers typically think of when examining risky choices but that are not applicable to our setting because of features of the policy change. After having ruled out alternative mechanisms, we then conduct an online laboratory experiment that captures the key features of the regrade decision and allows us to directly elicit participants' risk tolerance, beliefs, and confidence.

4.3. Alternative Explanations

In addition to the two mechanisms discussed above, there are several other mechanisms motivated by prior

Table 4. Likelihood of Submitting a Subsequent Regrade Conditional on First Regrade Result by Gender

	All	Post	Written	Written post
<i>Intercept</i>	0.097*** 0.003	0.108*** 0.004	0.090*** 0.003	0.105*** 0.005
<i>Female</i>	-0.023*** 0.004	-0.028*** 0.005	-0.021*** 0.005	-0.024*** 0.007
<i>In Favor</i>	0.025*** 0.006	0.021** 0.007	0.032*** 0.007	0.030*** 0.009
<i>Against</i>	-0.024* 0.011	-0.033** 0.012	-0.005 0.013	-0.017 0.014
<i>Female × In Favor</i>	0.001 0.008	0.003 0.009	0.000 0.010	-0.003 0.012
<i>Female × Against</i>	0.016 0.015	0.021 0.017	0.001 0.020	0.005 0.022
Observations	56,998	39,569	36,638	23,808

Notes. Impact of a student's first regrade outcome on their likelihood of submitting for regrades in the future. We estimate

$$Y_i = \alpha_0 + \beta_1 \text{Female} + \beta_2 \text{InFavor} + \beta_3 \text{Against} + \beta_4 (\text{Female} * \text{InFavor}) + \beta_5 (\text{Female} * \text{Against}) + \epsilon_i$$

We consider this outcome a proxy for belief updating about the likelihood of future success. Column (1) presents the estimates for all regrades submitted, column (2) presents the estimates for just post period regrades, column (3) presents the estimates for all written exams, and column (4) presents the estimates for written exams in the post period. Because of the low number of grades that are decreased in the period before the policy change, we do not use the difference in differences specification here. Standard errors are presented below the estimates.

⁺ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

work that might be at play: differences in perceived cost of requesting a regrade, differences in ambiguity aversion, differences in attention, and differences in how they value grades. However, they are not likely driving the results in our setting. We explain why below.

First, men and women may differ in the costs that they assign to making requests (Li and Zafar 2023, Roussille 2024). This is most likely to be the case in social contexts, where backlash is a more plausible concern (Bowles et al. 2007). The regrade request system in Norway minimizes these costs because justifications are typically requested and given in written form, and regrade requests are anonymous and not reported to the original grader. Student anonymity at the time of grading also makes direct and anticipated discrimination unlikely in our setting (Altonji and Blank 1999, Goldin and Rouse 2000, Babcock et al. 2017, Bohren et al. 2019). This stands in contrast to the U.S. system, where regrade requests typically require a direct interaction with a professor or teaching assistant (Li and Zafar 2023). Because the regrade system remains unchanged from the student perspective, the reform we study does not affect the costs. As a result, we consider it less likely that the differences we observe are attributable to the costs of asking. Additionally, we expect such costs to be independent of the original grade. As such, an explanation relying on gender differences in costs would predict a gender gap in regrade requests for F grades, which we do not observe.

Second, differential ambiguity aversion could also be at play in the regrade setting (Borghans et al. 2009, l'Haridon et al. 2018, Coffman et al. 2024). However, the policy change we study, although changing the underlying probabilities, does not affect students' metaknowledge about the underlying probabilities. As a result, it is unlikely that gender differences in ambiguity aversion are driving our results.

Third, the change in regrade policy may simply have directed attention to regrades, causing proportional increases across all grade levels in regrade requests. We rule this out by looking at B grades (Table 2). The expected value of asking for a regrade for B grades becomes negative after the policy change. Consistent with the change in expected value, B grades are the only grades that do not significantly increase in request likelihood after the policy change. This pattern is consistent with a responsiveness to the risks of asking, and not a uniform increase in regrade requests in response to the policy.

Finally, recent evidence suggests that men and women may value grades differently. Specifically, women appear to assign a considerably higher value to a one grade improvement than men (Ugalde 2024). If women value grades more than men, then the increased variance of a regrade would be less attractive to them in the post

period. This is particularly true if women are more loss averse than men. However, this mechanism would imply that women would seek more regrades for F grades relative to men, both before and after the policy. We do not observe this in the data.

Taken together, our analyses suggest that gender differences in risk preferences or loss aversion are not sufficient to explain the increased gender gap after the policy change. Features of our setting also minimize the role of other channels like gender differences in the valuation of grades, ambiguity aversion, and costs of asking. Gender differences in confidence or beliefs, therefore, seem to be necessary to rationalize the data and more broadly consistent with the patterns we observe. We also directly observe evidence of learning from prior experience, which is consistent with an important belief-based channel for deciding whether to regrade, and suggestive but weak evidence that women have more pessimistic priors. To examine these mechanisms further, we conduct an online laboratory experiment that captures the key features of the regrade decision and allows us to directly elicit participants' risk tolerance, beliefs, and confidence.

5. Experimental Results

We design an experiment to examine the role risk preferences and belief-based mechanisms play in driving willingness to ask for regrades. The experiment allows us to directly elicit a proxy for risk preferences used in the literature (Dohmen et al. 2011), as well as beliefs about the likelihoods of regrade outcomes and beliefs about performance. These data, which we do not have in the administrative data, can help us estimate how much different factors contribute to peoples' demand for regrades. The experiment also allows us to directly manipulate the presence of downside risk to examine what role, if any, downside risk plays in driving the gender gap. Finally, because the experiment is conducted online in a U.S. sample of participants, it allows us to examine whether the effect we find in NHH students is a more general finding.

Participants on Prolific ($N = 972$) completed a timed counting 1s task with 10 tables.¹¹ For each table scored correctly, participants earned 200 experimental points, which translated to 20 cents. Participants were informed that a computer would score their work but could make mistakes with an unknown probability. The computer mistakes could benefit or hurt the participants' score. Participants were also informed that they could pay for a regrade. In one condition, participants who bought a regrade had both upside and downside risk. That is, their score could be improved, reduced, or left the same. In the other condition, participants could only have their scores improved or left the same (i.e., mistakes in their favor would not be corrected).

After completing the task, we asked for participants' performance expectations. These expectations were incentivized. Specifically, we asked "How many tables out of 10 do you think you answered correctly? If your answer is correct, you will receive a bonus of 10 cents. This means it is in your interest to report your best guess." We then informed the participants of the computer score and elicited willingness to pay (WTP) for a regrade using a slider that ranged from 0 to 200 experimental points. We then elicited participants beliefs about the likelihood that their score would improve, stay the same, or decrease if they received a regrade. The beliefs were required to sum to 100. We did not incentivize these beliefs. After they found out the result of the regrade, we elicited a set of demographics including gender and overall risk tolerance on a zero to five scale (Dohmen et al. 2011, Buser et al. 2021). The exact working of this question was "Are you generally a person who is fully prepared to take risks or do you to try to avoid taking risks? Please select a number from 0 to 5." The zero item was labeled as "not at all willing to take risks (0)" and the five item was labeled "very willing to take risks (5)."

Participants completed the study with a median time of 8 minutes and 36 seconds. They were compensated with a base payment of \$1. Participants received an average bonus of approximately 87 cents. The average age was 42.36 years. All participants were currently living in the United States, and 93.7% had English as their first language. Additional summary statistics from the experiment data are provided in Panel B of Table A1 in the Online Appendix.

Our manipulation of perceptions of upside and downside risk are successful. On average, participants in the upside and downside condition believe that they will see their score increase with 35.36% and decrease with 28.31%. Participants in the upside only condition are more optimistic about the likelihood of a regrade increasing their score ($\beta_{just\ upside} = 3.28$, standard error (SE) = 1.87) and substantially less likely to think that their score will decrease ($\beta_{just\ upside} = -18.30$, SE = 1.42).

Consistent with prior work, we observe gender gaps in risk preferences, overconfidence, and beliefs. Women report being 0.67 scale points less willing to take risk (SE = 0.084) out of 5. Although men and women report statistically similar performance expectations ($\beta_{overall} = -0.18$, SE = 0.16), men are more overconfident, deviating from their true score by 0.71 answers (SE = 0.10), but women only overestimate by 0.29 (SE = 0.10). Men are also more optimistic than women about upside risks in both of our risk conditions ($\beta_{upside+downside} = 5.37$, SE = 2.55; $\beta_{just\ upside} = 6.85$, SE = 2.72), but women are only more pessimistic about downside risk when it is relevant ($\beta_{upside+downside} = 4.02$, SE = 2.29; $\beta_{just\ upside} = -1.01$, SE

= 1.74). For a comparison of the gender gaps between conditions see Table A17 in the Online Appendix.

We observe a substantial and significant gender gap of 8.07 experimental points (23.8% of male WTP) when participants are exposed to both upside and downside risk (Table 5). There is no significant gender gap when there is no downside risk (0.7 experimental points, 2% of male WTP). Although we cannot statistically reject that these gender gaps are the same, the difference is very large (7.38 experimental points, SE = 5.78) and consistent with the empirical results obtained when evaluating the grading policy at NHH. Table 5 examines the relationship within each condition of WTP with risk preferences, beliefs about regrade outcomes, and overconfidence about performance. We observe that WTP in both conditions is significantly correlated with risk preferences, beliefs about the likelihood of regrade outcomes, and overconfidence about performance in pairwise examinations. When including all variables in the same regression, beliefs and overconfidence in performance are significantly correlated with WTP; a result that aligns closely with our findings from the administrative data analysis that belief-based mechanisms appear to be more important than preference based ones.¹² Additionally, the gender gap in the upside and downside risk condition shrinks by two thirds when accounting for risk preferences, beliefs, and overconfidence about performance and is no longer significant (Table 5, Panel A, column (6)). To facilitate comparison of magnitudes we examine these correlations in standardized terms (Table A18 in the online table). In the full sample, a 1-SD increase in beliefs about the probability of increase corresponds to an 18.07-point increase in WTP (SE = 1.66). A similar 1-SD increase in overconfidence corresponds to a 4.36-point increase in WTP (SE = 1.54). A 1-SD increase in risk tolerance corresponds to a 3.1-point increase in WTP (SE = 1.39).

Our willingness-to-pay variable is skewed but also contains many zeros. Recent work suggests several ways to address this issue (Chen and Roth 2024, Mullahy and Norton 2024). To this end, we also present results from a Poisson regression (Table A19 in the Online Appendix), modeling the extensive and intensive margins separately (Table A20 in the Online Appendix), and as a two-part model in which we first estimate the probability of having a positive willingness-to-pay value and then use those probabilities as weights in a probit model that examines the effect on willingness-to-pay conditional on having a positive value (Mullahy and Norton 2024) (Table A21 in the Online Appendix). The results from these alternative specifications are similar to our main estimates. Additionally, they suggest that the gender gap we observe is driven mostly by differences in the intensive margin.

Table 5. Experimental Results by Condition

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Panel A: Experimental results the condition with both upside and downside risk						
(Intercept)	33.635*** (2.946)	14.629* (6.540)	7.109* (3.412)	39.715*** (3.804)	30.174*** (2.853)	0.069 (6.038)
Female	-7.831* (3.969)	-4.167 (4.100)	-4.072 (3.608)	-6.905+ (3.980)	-5.920 (3.906)	-2.318 (3.776)
Risk		4.973** (1.586)				1.557 (1.512)
Up Beliefs			0.699*** (0.089)			0.653*** (0.090)
Down Beliefs				-0.231** (0.071)		0.038 (0.048)
Overconfidence					4.508*** (1.137)	2.375* (1.046)
Observations	465	465	465	465	465	465
Panel B: Experimental results the condition with only upside risk						
(Intercept)	33.322*** (2.895)	17.646** (6.005)	6.411* (2.892)	34.455*** (3.322)	30.556*** (2.685)	-4.767 (6.269)
Female	-0.561 (4.208)	1.953 (4.279)	3.812 (3.822)	-0.670 (4.230)	1.169 (4.116)	6.050 (3.999)
Risk		4.143** (1.561)				2.760+ (1.420)
Up Beliefs			0.638*** (0.073)			0.605*** (0.074)
Down Beliefs				-0.108 (0.117)		0.097 (0.111)
Overconfidence					4.210*** (0.950)	1.681+ (0.919)
Observations	505	505	505	505	505	505

Notes. Willingness to pay results for each experimental condition and the correlation between willingness to pay and other variables of interest. Panel A shows the results for the condition with both upside and downside risk, and Panel B shows the results for the condition with only upside risk. Model 1 includes a binary variable for Female, which shows the gap in willingness to pay between women and men, Model 2 includes Female and risk tolerance on a six-point scale, Model 3 includes Female and beliefs about upside probabilities of regrades on a 0–100 scale, Model 4 includes Female and beliefs about downside probabilities of regrades on a 0–100 scale, Model 5 includes Female and overconfidence, which we operationalize as the gap between expected and actual performance, and Model 6 includes all variables. Robust standard errors are included in parentheses. There is a significant gender gap in the condition with downside risk, but no significant gender gap in the condition with only upside risk. The gender gap in the condition with downside risk is reduced by two thirds when controlling for other variables of interest.

+ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Taken together, the results from our online experiment suggest that the effect we document in a real-world choice environment extends to an online laboratory environment. The online experiment helps us overcome a data limitation in our administrative data by eliciting a proxy for risk preferences, beliefs about regrade outcomes, and beliefs about performance directly. We replicate findings in prior work that men are more risk tolerant, more optimistic, and more confident about their performance. Consistent with our findings in the post period of our administrative data, men are more willing to pay for a regrade when downside risk is present than women. We present suggestive evidence that this gap narrows when scores can only be improved. We also find that beliefs and overconfidence in performance primarily contribute to the WTP for regrades and explain a large share of the gender gap in the upside and downside condition.

6. Discussion

Do changes in the riskiness of choices affect gender gaps in outcomes? In this paper, we isolate the causal effect of changing the riskiness of choices (altering the probabilities of the outcomes associated with choices) on the gender gap in real world choices and provide strong evidence of the mechanisms through which the effects operate.

The policy we study was implemented in the second half of 2014 and altered the procedure for grading Norwegian university students' request regrades on assignments and exams. By exploiting unique individual-level administrative data from Norway's leading business school on all student assignments, grades, and regrades (requests and results), we show that both the expected value of a regrade request, as well as the likelihood of negative outcomes, increased substantially as a result of the policy. Because the reform narrowly impacts the

probabilities associated with the different possible regrade outcomes, while leaving all other elements of the regrade process unchanged, it provides an ideal setting for examining how policies that shift the riskiness of choices impact gender gaps in real world decisions.

Using a DiD framework, we compare the behavior of men and women pre and post the policy change and provide three sets of results. First, we document a substantial increase in the gender gap in regrade requests in response to the policy. Although both men and women became more likely to request a regrade after the policy change, men responded much more strongly to the reform. The gender gap in regrade requests increased by 90% after the policy was implemented.

Second, the policy-induced gender gap in regrade requests translates into a gender gap in important downstream outcomes: grade points earned via regrade. This implies that the ostensibly gender-neutral regrade policy may augment gender differences in labor market outcomes and highlights the role that social institutions play in shaping the dynamics of gender gaps in society.

Finally, we show that the expansion of the gender gap is likely driven by gender differences in beliefs and confidence. In terms of our mechanism analyses, it is always challenging to rely on subsample analyses as a means to disentangle mechanisms (because of the possibility that several mechanisms are at play in any one such analysis). However, in our case, we examine mechanisms through a variety of different methods, including structural estimations, heterogeneity analyses, and online experiments. All these results point to the same mechanism, gender differences in beliefs and confidence, being the most important. The consistency of our findings across these different analyses provide strong support for the results we find. The sum total of these several analyses strongly suggest that a belief mechanism is the most consistent with the data and that a risk preference-based mechanism is not sufficient to produce the gender gaps we document. We also emphasize a direct test for the role of beliefs in the decision to request a regrade, specifically that students appear to learn from prior regrade outcomes. We generalize our findings using an online experiment and show a substantial gender gap when the possibility of downside risk is present, but not when it is absent.

A major advantage of our setting is the limited scope for mechanisms other than changes in the riskiness of choices to drive our findings. For example, discrimination (Altonji and Blank 1999, Goldin and Rouse 2000, Bohren et al. 2019) and biased expectations of evaluators (Babcock et al. 2017) are unlikely to drive the findings since students know the actual grading is anonymous. As a result, our setting provides an ideal test of how changes in the riskiness of choices affect gender gaps in outcomes.

The key contribution of this paper is to show how ostensibly neutral policies can generate gaps across groups who differ in their risk-taking behavior. Many policy reforms affect the riskiness of choices, and knowing how changes in the riskiness of choices interact with group-specific risk preferences and beliefs to generate differences in program take-up and inequalities across groups is crucial for understanding how policy impacts efficiency and equity in society (Currie 2004, Heckman and Landersø 2022).

Specifically, the paper demonstrates that failing to account for systematic differences in risk taking behavior across groups when making policies can have detrimental effects on certain groups. If policy changes alter outcome probabilities of choices, particularly those with negative outcomes, the impact will likely vary across people due to the type of selection we document. In our setting, because the expected value is positive, the increase in asking driven by the policy change disproportionately benefits men. In another setting, for example where the expected value of asking was actually negative, men might ask too much and pay an increased cost relative to women. This is also true in our setting. The expected value of requesting a regrade with an original grade of B decreases after the policy change. We see that women lose fewer grade points than men from B grades in the post period of our study. That is, men still ask more when the outcomes have a negative expected value, but it becomes harmful to them.

More broadly, our paper raises the question of how systematic differences across groups in fundamental economic constructs like risk preferences and beliefs interact with policy structures to influence the use of opportunities. We show that even in the absence of meaningful discrimination, the structure of the decision environment can widen gaps in terms of who takes advantage of services and opportunities.

When interpreting our results, it is also important to note that it is not clear what the optimal tradeoff between ex ante fairness and ex post equity is. In terms of ex ante fairness, the goal of the national regrading reform was to provide students with a fair chance of reassessment that was not anchored by prior results. Although there was a considerable public debate on the question of whether to blind the regrade process or not, the National Student Organization of Norway was a strong proponent of the policy. Thus, the reform had strong a priori support from the community and arguably generated an overall welfare improvement among the students. The negative impact on inequality between men and women was an unanticipated implication of this reform, and these policy tradeoffs are hard to predict prior to policy implementation. Our results highlight the fact that a tradeoff exists between ex ante fairness and ex post equity in this context, but

resolving that tradeoff depends on how strongly society weights those two values.

Acknowledgments

The authors thank Ingar Haaland, Alex Imas, Katrine Løken, Bertil Tungodden, Barton Willage, and Basit Zafar for helpful comments and seminar participants at Carnegie Mellon University, the Rady School of Management, the Early Career Behavioral Economics Conference, Helsinki Graduate School of Economics, and the Society for Judgment and Decision Making Conference.

Endnotes

¹ Although the exact shift in regrade outcomes was unknown at implementation, the student organizations believed the policy would generate a favorable outcome for their members (<https://www.idunn.no/doi/10.18261/issn.1893-8981-2019-04-02>).

² Specifically, we show that both men and women who originally received F grades respond to the new policy by increasing regrade requests. However, there is no significant gender difference in regrade requests for Fs either before or after the policy change. This stands in contrast to what we find for other grades, in which there is a large increase in downside risk and a subsequent increase in the gender gap in regrade requests.

³ It would be interesting to examine what percentage of those who ask for an explanation (and receive an explanation) request a regrade. Unfortunately, the exam office does not record this information in the administrative registers. However, to the extent that gender differences emerge already in the initial justification stage, our results should be considered lower bounds of the gender gap in the entire reassessment process.

⁴ For additional information see <https://studentorget.no/index.php?artikelid=13762&expand=37> or <https://www.jus.uio.no/studier/aktuelle-saker/2016/bilder/kd-notat-blind-klagesensur.pdf>.

⁵ Students receiving As have no reason to request a regrade; students receiving Fs have no downside risk associated with requesting a regrade.

⁶ Additional robustness and sensitivity analyses are provided in the Online Appendix. Specifically, in the Online Appendix, Figure A2 shows the simple difference in differences interaction with different sample restrictions, Table A2 shows the results with Class X Year fixed effects, and Table A3 shows our main result across each of the departments in the bachelors' program.

⁷ For full regression output see Table A4 in the Online Appendix.

⁸ In the Online Appendix, we have also examined heterogeneity by the level of education (bachelor or master) of the students. It is possible that the expected value of regrade request changed differentially across the undergraduate and the graduate level, and as a consequence, that students at those levels responded differently to the policy. In Figure A4, we show that the effect of the policy on the expected value of a regrade request was substantially larger at the undergraduate level, and in Figure A5, we show that the policy increased the gender gap in regrade requests much more at the undergraduate level.

⁹ Receiving a downgrade to an F is disproportionately bad, and men are more likely to receive a downgrade after the reform. If this includes an increased likelihood of receiving an F, the effect of the reform on expected utility could therefore be better for women than for men even though grade increases are more likely for men. However, as shown in Table A9 in the Online Appendix, although there is a slight reduction in the overall probability of receiving an F as a

consequence of the blinding policy, there is no gender difference in the impact of the policy on the likelihood of receiving an F.

¹⁰ One important question related to this analysis concerns the frequency through which students receive an F via a regrade. To this end, in the Online Appendix, Table A11 shows the distribution of new grades before and after the reform by original grade, Table A12 shows the likelihood of receiving an F grade by original grade, and Figure A7 shows the distribution of regrade outcomes for each original grade. Taken together, these results show that regrade outcomes of F are fairly likely for original E grades (9.3% of requests), substantially less common for original D grades (0.007%) and basically do not happen otherwise. The figure illustrates that the majority of regrades fall within one letter grade of the original grade but that there are nontrivial shares of students that receive grades that differ with more than one letter grade from the original grade as well.

¹¹ Data collection and analysis were preregistered using asPredicted (#150550). For additional analyses see Tables A14, A15, and A16 in the Online Appendix.

¹² It is worth noting that beliefs about the upside and downside of asking for a regrade are significantly correlated ($\beta = -0.32$, $SE = 0.038$), which may explain why upside beliefs are significant while downside beliefs are not statistically significant and flip sign relative to column (4).

References

- Almås I, Cappelen AW, Salvanes KG, Sørensen EØ, Tungodden B (2016) What explains the gender gap in college track dropout? Experimental and administrative evidence. *Amer. Econom. Rev.* 106(5):296–302.
- Altonji JG, Blank RM (1999) Race and gender in the labor market. Ashenfelter OC, Card D, eds. *Handbook of Labor Economics* (Elsevier, Amsterdam), 3143–3259.
- Babcock L, Laschevar S (2003) *Women Don't Ask: Negotiation and the Gender Divide* (Princeton University Press, Princeton, NJ).
- Babcock L, Laschevar S, Recalde MP, Vesterlund L, Weingart L (2017) Gender differences in accepting and receiving requests for tasks with low promotability. *Amer. Econom. Rev.* 107(3):714–747.
- Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* 335(6068):579–582.
- Baldiga K (2014) Gender differences in willingness to guess. *Management Sci.* 60(2):434–448.
- Barber BM, Odean T (2001) Boys will be boys: Gender, overconfidence, and common stock investment. *Quart. J. Econom.* 116(1): 261–292.
- Bertrand M (2011) New perspectives on gender. *Handbook of Labor Economics*, vol. 4 (Elsevier, New York), 1543–1590.
- Blau FD, Kahn LM (2017) The gender wage gap: Extent, trends, and explanations. *J. Econom. Literature* 55(3):789–865.
- Bohren JA, Imas A, Rosenberg M (2019) The dynamics of discrimination: Theory and evidence. *Amer. Econom. Rev.* 109(10):3395–3436.
- Borghans L, Heckman JJ, Golsteyn BHH, Meijers H (2009) Gender differences in risk aversion and ambiguity aversion. *J. Eur. Econom. Assoc.* 7(2–3):649–658.
- Bowles HR, Babcock L, Lai L (2007) Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organ. Behav. Human Decision Processes* 103(1):84–103.
- Buser T, Cappelen AW, Tungodden B (2021) Fairness and willingness to compete. Preprint, submitted March 12, <http://dx.doi.org/10.2139/ssrn.3804698>.
- Buser T, Niederle M, Oosterbeek H (2014) Gender, competitiveness, and career choices. *Quart. J. Econom.* 129(3):1409–1447.

- Card D, Heining J, Kline P (2013) Workplace heterogeneity and the rise of West German wage inequality. *Quart. J. Econom.* 128(3): 967–1015.
- Charness G, Gneezy U (2012) Strong evidence for gender differences in risk taking. *J. Econom. Behav. Organ.* 83(1):50–58.
- Chen J, Roth J (2024) Logs with zeros? Some problems and solutions. *Quart. J. Econom.* 139(2):891–936.
- Coffman KB, Klinowski D (2020) The impact of penalties for wrong answers on the gender gap in test scores. *Proc. Natl. Acad. Sci. USA* 117(16):8794–8803.
- Coffman K, Collis M, Kulkarni L (2021) *Stereotypes and Belief Updating* (Harvard Business School, Cambridge, MA).
- Coffman KB, Collis MR, Kulkarni L (2024) Whether to apply. *Management Sci.* 70(7):4649–4669.
- Cortés P, Pan J, Pilossoph L, Reuben E, Zafar B (2023) Gender differences in job search and the earnings gap: Evidence from the field and lab. *Quart. J. Econom.* 138(4):2069–2126.
- Crosen R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Literature* 47(2):448–474.
- Currie J (2004) The take up of social benefits. NBER Working Paper No. 10488, National Bureau of Economic Research, Cambridge, MA.
- Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econom. Assoc.* 9(3):522–550.
- Exley CL, Kessler JB (2022) The gender gap in self-promotion. *Quart. J. Econom.* 137(3):1345–1381.
- Exley CL, Nielsen K (2024) The gender gap in confidence: Expected but not accounted for. *Amer. Econom. Rev.* 114(3):851–885.
- Falk A, Becker A, Dohmen T, Enke B, Huffman D, Sunde U (2018) Global evidence on economic preferences. *Quart. J. Econom.* 133(4): 1645–1692.
- Filippin A, Crosetto P (2016) A reconsideration of gender differences in risk attitudes. *Management Sci.* 62(11):3138–3160.
- Flory JA, Leibbrandt A, List JA (2015) Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions. *Rev. Econom. Stud.* 82(1):122–155.
- Floyd E, Tomar S, Lee DJ (2024) Making the grade (but not disclosing it): How withholding grades affects student behavior and employment. *Management Sci.* 70(4):2497–2517.
- Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118(3): 1049–1074.
- Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of “blind” auditions on female musicians. *Amer. Econom. Rev.* 90(4):715–741.
- Heckman J, Landersø R (2022) Lessons for Americans from Denmark about inequality and social mobility. *Labour Econom.* 77:101999.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
- Kessler JB, Low C, Sullivan CD (2019) Incentivized resume rating: Eliciting employer preferences without deception. *Amer. Econom. Rev.* 109(11):3713–3744.
- l’Haridon O, Vieider FM, Aycinena D, Bandur A, Belianin A, Cingl L, Kithiyal A, et al. (2018) Off the charts: Massive unexplained heterogeneity in a global study of ambiguity attitudes. *Rev. Econom. Statist.* 100(4):664–677.
- Landaud F, Maurin É, Willage B, Willén A (2024) The value of a high school GPA. *Rev. Econom. Statist.* Forthcoming.
- Leibbrandt A, List JA (2015) Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Sci.* 61(9):2016–2024.
- Li CH, Zafar B (2023) Ask and you shall receive? Gender differences in regrades in college. *Amer. Econom. J. Econom. Policy* 15(2): 359–394.
- Meier S, Sprenger C (2010) Present-biased preferences and credit card borrowing. *Amer. Econom. J. Appl. Econom.* 2(1):193–210.
- Mullahy J, Norton EC (2024) Why transform y? The pitfalls of transformed regressions with a mass at zero. *Oxford Bull. Econom. Statist.* 86(2):417–447.
- Murciano-Goroff R (2022) Missing women in tech: The labor market for highly skilled software engineers. *Management Sci.* 68(5): 3262–3281.
- Niederle M (2015) *Gender* (Princeton University Press, Princeton NJ).
- Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122(3): 1067–1101.
- Olivetti C, Petrongolo B (2016) The evolution of gender gaps in industrialized countries. *Annu. Rev. Econom.* 8:405–434.
- Quadlin N (2018) The mark of a woman’s record: Gender and academic performance in hiring. *Amer. Sociol. Rev.* 83(2):331–360.
- Riise J, Willage B, Willén A (2022) Can female doctors cure the gender STEM gap? Evidence from exogenously-assigned general practitioners. *Rev. Econom. Statist.* 104(4):621–635.
- Roussille N (2024) The role of the ask gap in gender pay inequality. *Quart. J. Econom.* 139(3):1557–1610.
- Samek A (2019) Gender differences in job entry decisions: A university-wide field experiment. *Management Sci.* 65(7):3272–3281.
- Schroyen F, Aarbu KO (2018) Attitudes towards large income risk in welfare states: An international comparison. *Economica* 85(340): 846–872.
- Small DA, Gelfand M, Babcock L, Gettman H (2007) Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation. *J. Personality Soc. Psych.* 93(4):600.
- Sutter M, Glätzle-Rützler D (2015) Gender differences in the willingness to compete emerge early in life and persist. *Management Sci.* 61(10):2339–23354.
- Sutter M, Kocher MG, Glätzle-Rützler D, Trautmann ST (2013) Impatience and uncertainty: Experimental decisions predict adolescents’ field behavior. *Amer. Econom. Rev.* 103(1):510–531.
- Thöni C, Volk S (2021) Converging evidence for greater male variability in time, risk, and social preferences. *Proc. Natl. Acad. Sci. USA* 118(23):e2026112118.
- Tungodden J, Willén A (2023) When parents decide: Gender differences in competitiveness. *J. Political Econom.* 131(3):751–801.
- Ugalde MP (2024) Gender, grade sensitivity, and major choice. https://pao1607.github.io/paolaugaldea.com/GradesGender_draft.pdf.
- von Gaudecker H-M, Van Soest A, Wengström E (2011) Heterogeneity in risky choice behavior in a broad population. *Amer. Econom. Rev.* 101(2):664–694.