Essays in Applied Microeconomics

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Yana Radeva

aus Sofia

2026

Dekan:	Prof. Dr. Martin Böse
Erstreferent:	Prof. Dr. Thomas Dohmen
Zweitreferent:	Prof. Dr. Pia Pinger
Tag der mündlichen Prüfung:	17. Juli 2026

Acknowledgments

This dissertation marks the culmination of a long journey, and it would not have been possible without the support and guidance of many wonderful people.

First and foremost, I am deeply grateful to my family. To my husband, Ruben, thank you for being an unfailing source of love, strength, and encouragement, and for standing by me through both the difficult moments and the joyful ones. To my mother, thank you for making this journey possible through your constant support, both for me and for my family. To Leo and Max, thank you for keeping me grounded and reminding me every day of what truly matters.

I would also like to express my sincere gratitude to my supervisors, Thomas Dohmen and Pia Pinger, for their expertise, insightful feedback, guidance, and reassurance throughout this process. Thomas has been a wonderful example of intellectual rigor, hard work, and dedication to research. Pia has been a true inspiration to me and one of the reasons I decided to pursue a PhD in the first place; she has been my role model throughout this journey.

I am especially grateful to my co-authors: Malin Siemers, Anna Person, Si Chen, Thomas Dohmen, Anna Schulze Tilling, and Elena Shvartsman. I am also grateful to my colleagues and fellow researchers in Bonn: Barbara Boelmann, Teodora Boneva, Antonia Entorf, Holger Gerhardt, Paul Grass, Radost Holler, Hannah Illing, Botond Kőszegi, Jan Scherer, Amelie Schiprowski, Philipp Schirmer, Lenard Simon, Jakob Wegmann, and Florian Zimmermann. Together, they shaped this dissertation through collaboration and discussion. I would also like to thank Simone Jost and Revena Penning for their help and support throughout this journey.

Finally, I would like to acknowledge everyone who contributed directly or indirectly to the realization of this thesis. I made it through mathematics and theoretical microeconomics thanks in no small part to Jan and Miriam. I am grateful to Radost for always being there, especially in her role as student representative, and for taking on both the large and the small battles. I thank Anna for the many hours we shared in the lab and Malin for the long lunches, coffee breaks, and the fruitful discussions.

Yana Radeva
Bonn, April 2026

Contents

List of Figures

List of Tables

Introduction

Applied microeconomics seeks to explain how economically relevant choices are shaped by incentives, institutions, beliefs, and the environments in which decisions are made. A central challenge in this endeavor is empirical: many determinants of behavior are correlated with unobserved individual characteristics, making causal inference difficult. This dissertation addresses that challenge in three settings. Across the chapters, I study how information, institutional settings, and acute psychological conditions affect policy preferences, occupational sorting, and risk-taking. While the substantive domains differ, all three chapters share a common objective: to identify causal determinants of choices.

The first two chapters speak directly to gender inequality in the labor market. A large body of literature shows that gender gaps in wages and occupations remain a central feature of modern economies (Goldin, 2024; Olivetti, Pan, and Petrongolo, 2024). Recent work further emphasizes the central role of children in shaping women's earnings trajectories (Kleven, Landais, and Søgaard, 2019; Cortés and Pan, 2023), while a parallel literature documents the importance of occupational sorting for the persistence of inequality (Goldin and Katz, 2016; Goldin et al., 2017). Together, these findings suggest that understanding inequality requires going beyond realized labor market outcomes and examining the beliefs, institutions, and moments of choice that generate them. The third chapter broadens this perspective by turning to a more general behavioral question: whether acute stress causally affects risky choice. It addresses the same underlying concern with how context shapes behavior and how credible empirical designs can distinguish genuine effects from spurious variation.

Chapter 1, *Children and the Gender Wage Gap: Beliefs, (Mis)perceptions, and Policy Preferences*, investigates whether beliefs about the sources of inequality shape support for policies designed to reduce inequality. Motivated by evidence that a substantial share of the remaining gender wage gap is related to the differential effect of children on the careers of women and men (Kleven, Landais, and Søgaard, 2019; Cortés and Pan, 2023; Goldin, 2024), we study whether the public recognizes the importance of this factor and whether such beliefs affect policy preferences. Using a survey experiment in the United States, we exogenously shift beliefs about the share of the gender wage gap explained by having children. The results show that respondents substantially underestimate the role of children in explaining the gender wage gap and that increasing the perceived importance of this channel reallocates support toward child-related policies and away from policies aimed at discrimination. In doing so, the chapter contributes to a broader literature on how beliefs about inequality affect policy preferences (Kuziemko et al., 2015; Alesina, Stantcheva, and Teso, 2018; Alesina, Ferroni, and Stantcheva, 2021; Settele, 2022; Haaland and Roth, 2023; Casarico, Schuetz, and Uebelmesser, 2024) by showing that beliefs about the causes of inequality, and not only its magnitude, matter for policy support.

Chapter 2, *Age at Labor Market Entry and Gender Occupational Segregation*, shifts attention from beliefs to institutions. We investigate whether the timing of labor market entry causally affects occupational sorting by gender. Occupational choice is a consequential and often persistent margin of adjustment, yet relatively little is known about how the age at which it is made shapes later career trajectories (Keane and Wolpin, 1997; Attanasio and Kaufmann, 2017). Identifying such an effect is difficult because it requires exogenous variation in age at decision-making. Exploiting reforms to school entry cutoff dates across German states, we use quasi-experimental variation in effective school entry and exit ages to identify the causal effect of entering the labor market younger. The findings show that men who enter the labor market at a younger age are more likely to sort into male-dominated occupations, whereas women's occupational choices are not significantly affected. We thus identify timing as a previously underappreciated determinant of gender occupational segregation and contribute both to the literature on occupational choice and to debates on early tracking and inequality (Gamoran and Mare, 1989; Slavin, 1990; Oakes, 2005).

Chapter 3, *Does Stress Impact Risk-Taking?*, turns from labor market outcomes to a more general question in economics: whether acute stress causally changes risky choice. This question matters because stress is pervasive in real-world decision environments, yet the existing evidence remains contradictory. Some studies find that stress increases risk-taking (Buckert et al., 2014; Bendahan et al., 2017), others find that it decreases risk-taking (Yamakawa et al., 2016; Cahlíková and Cingl, 2017), and still others report no effect (Sokol-Hessner et al., 2016; Veszteg et al., 2021; Parslow and Rose, 2022). We argue that this inconsistency may reflect both theoretical ambiguity and methodological limitations, in particular the reliance on small between-subject designs that are vulnerable to baseline heterogeneity in risk preferences. To address these concerns, we introduce a high-powered within-subject laboratory design in which each participant completes identical incentivized risky choices once in a *Control* condition and once in a biomarker-validated *Stress* condition. In contrast to much of the earlier literature, we find no evidence that acute stress causally affects risk-taking, neither on average nor through shifts toward or away from risk neutrality. The results suggest that at least part of the earlier mixed evidence may be attributable to sampling variation in baseline preferences rather than to genuine treatment effects.

The chapters are connected not only by their focus on economically relevant choices, but also by the close alignment between each research question and its empirical design. In Chapter 1, a survey experiment is the natural approach because beliefs about the causes of inequality must be shifted directly and exogenously to identify their effect on policy preferences. In Chapter 2, random assignment of labor market entry age is impossible, making institutional cutoff reforms a credible source of quasi-experimental variation. In Chapter 3, the treatment is temporary, physiologically complex, and costly to induce, so a controlled laboratory setting with repeated observation of the same individuals is especially valuable. The dissertation therefore brings together three complementary empirical strategies, each particularly well suited to the causal question at hand.

Combined, the three chapters foster our understanding of the causal forces that shape economically relevant choices. By combining experimental and quasi-experimental methods across these three distinct settings, this dissertation contributes to applied microeconomics both by shedding light on important determinants of policy preferences, occupational sorting, and risky choice, and by illustrating how distinct identification strategies can be matched to distinct causal questions in a disciplined and context-sensitive way.

References

Alesina, Alberto, Matteo F Ferroni, and Stefanie Stantcheva. 2021. "Perceptions of racial gaps, their causes, and ways to reduce them." Working Paper 29245. National Bureau of Economic Research. [2]

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational mobility and preferences for redistribution." *American Economic Review* 108 (2): 521–54. [2]

Attanasio, Orazio P, and Katja M Kaufmann. 2017. "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations." *Journal of Economic Behavior & Organization* 140: 35–55. [2]

Bendahan, Samuel, Lorenz Goette, J. Thoresen, Leyla Loued-Khenissi, Fiona Hollis, and Carmen Sandi. 2017. "Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk." *European Journal of Neuroscience* 45 (7): 877–85. [2]

Buckert, Magdalena, Christiane Schwieren, Brigitte M. Kudielka, and Christian J. Fiebach. 2014. "Acute stress affects risk taking but not ambiguity aversion." *Frontiers in Neuroscience* 8: 82. [2]

Cahlíková, Jana, and Lubomír Cingl. 2017. "Risk preferences under acute stress." *Experimental Economics* 20 (1): 209–36. [2]

Casarico, Alessandra, Jana Schuetz, and Silke Uebelmesser. 2024. "Gender inequality over the life cycle, information provision and policy preferences." Working Paper 10916. CESifo Working Paper Series. [2]

Cortés, Patricia, and Jessica Pan. 2023. "Children and the remaining gender gaps in the labor market." *Journal of Economic Literature* 61 (4): 1359–409. [1]

Gamoran, Adam, and Robert D Mare. 1989. "Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?" *American Journal of Sociology* 94 (5): 1146–83. [2]

Goldin, Claudia. 2024. "Nobel lecture: An evolving economic force." *American Economic Review* 114 (6): 1515–39. [1]

Goldin, Claudia, and Lawrence F. Katz. 2016. "A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation." *Journal of Labor Economics* 34 (3): 705–46. [1]

Goldin, Claudia, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth. 2017. "The expanding gender earnings gap: Evidence from the LEHD-2000 Census." *American Economic Review* 107 (5): 110–14. [1]

Haaland, Ingar, and Christopher Roth. 2023. "Beliefs about racial discrimination and support for pro-black policies." *Review of Economics and Statistics* 105 (1): 40–53. [2]

Keane, Michael P, and Kenneth I Wolpin. 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3): 473–522. [2]

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11 (4): 181–209. [1]

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How elastic are references for redistribution? Evidence from randomized survey experiments." *American Economic Review* 105 (4): 1478–508. [2]

Oakes, Jeannie. 2005. *Keeping track: How schools structure inequality.* Yale University Press. [2]

Olivetti, Claudia, Jessica Pan, and Barbara Petrongolo. 2024. "The evolution of gender in the labor market." In *Handbook of Labor Economics,* 5: 619–77. Elsevier. [1]

Parslow, Elle, and Julia Rose. 2022. "Stress and risk – Preferences versus noise." *Judgment and Decision Making* 17 (4): 883–936. [2]

Settele, Sonja. 2022. “How do beliefs about the gender wage gap affect the demand for public policy?” *American Economic Journal: Economic Policy* 14 (2): 475–508. [2]

Slavin, Robert E. 1990. “Achievement effects of ability grouping in secondary schools: A best-evidence synthesis.” *Review of Educational Research* 60 (3): 471–99. [2]

Sokol-Hessner, Peter, Candace M. Raio, Sarah P. Gottesman, Sandra F. Lackovic, and Elizabeth A. Phelps. 2016. “Acute stress does not affect risky monetary decision-making.” *Neurobiology of Stress* 5: 19–25. [2]

Veszteg, Róbert F., Kaori Yamakawa, Tetsuya Matsubayashi, and Michiko Ueda. 2021. “Acute stress does not affect economic behavior in the experimental laboratory.” *PLoS One* 16 (1): e0244881. [2]

Yamakawa, Kaori, Hideki Ohira, Masahiro Matsunaga, and Tokiko Isowa. 2016. “Prolonged effects of acute stress on decision-making under risk: A human psychophysiological study.” *Frontiers in Human Neuroscience* 10: 444. [2]

Chapter 1

Children and the Gender Wage Gap: Beliefs, (Mis)perceptions, and Policy Preferences*

Joint with Malin Siemers

1.1 Introduction

Despite substantial progress over recent decades, a persistent gender wage gap remains (see e.g., Olivetti, Pan, and Petrongolo, 2024). Recent research shows that a substantial portion of this remaining gap is driven by the differential impact of children on the earnings trajectories of men and women (e.g., Kleven, Landais, and Søgaard, 2019; Cortés and Pan, 2023). This evidence has important implications for the effectiveness of different policies aimed at reducing gender inequality (see e.g., Goldin, 2024). Implementation of such policies depends, in part, on public awareness of these mechanisms and support for interventions, especially those targeting the child-related part of the gap. This raises the question of whether people's beliefs about the role of children in the gender wage gap align with the academic evidence and how such beliefs shape policy preferences.

* We are grateful to Teodora Boneva, Thomas Dohmen, Paul Grass, Hannah Illing, Botond Kőszegi, Valentina Melentyeva, Jessica Pan, Barbara Petrongolo, Pia Pinger, Amelie Schiprowski, Philipp Schirmer, Sonkurt Şen, Sonja Settele, and Florian Zimmermann for valuable comments and support. We also thank the participants at the ECONtribute Junior Workshop in Gender and Labor Economics and the Bonn Applied Micro Coffee for helpful feedback and suggestions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866 and through the Collaborative Research Center Transregio 224 (CRC TR 224 - project A01 & A02). The study received ethics approval from the German Association for Experimental Economic Research e.V. (no. yRI4L6PQ). The study was preregistered on AsPredicted (no. 230,372).

Individuals may hold systematically different beliefs about the causes of the gender wage gap, particularly regarding the contribution of having children. Such beliefs may shape perceptions of which policies are likely to be effective, thereby influencing the types of policies individuals support. However, it remains an open question whether beliefs about the importance of specific factors, such as child-related pay differences, translate into greater support for policies targeting those factors, potentially at the expense of support for policies addressing other drivers of the gap.

Differences in beliefs about the underlying causes of the gender wage gap may not only shape support for specific policies, but also influence overall support for reducing the gap. If different drivers are perceived as differing in fairness or efficiency, these perceptions may translate into differences in general support for policy intervention. Such variation in beliefs may therefore help explain the considerable disagreement about whether and how strongly policy should aim to address the remaining disparity. Yet systematic evidence on whether and how beliefs about the causes of the gender wage gap, particularly the role of children, affect policy preferences remains limited.

In this paper, we examine whether beliefs about the contribution of children to the gender wage gap align with the academic evidence and investigate whether and how these beliefs affect overall support for reducing the gap, support for specific policies, and how they shift preferences between policies aimed at mitigating child-related pay differences and those targeting other factors.

To address these questions, we design and conduct an online survey experiment in the U.S. The experiment implements a 2x2 between-subject design that sequentially varies, first, the assumptions participants are asked to make about the effect of having children on men's earnings and, second, whether they receive information about the share of the gender wage gap explained by having children. Our experimental design is explicitly aimed at identifying the causal effect of beliefs about the role of children in explaining the gender wage gap on policy preferences. Studying this relationship using observational data alone is inherently challenging, as individuals' beliefs about the sources of the gender wage gap are likely correlated with unobserved characteristics, such as political ideology or selective exposure to information, which complicates causal inference. An experimental setting that allows for exogenous variation in beliefs, precise measurement of policy preferences and underlying perceptions, and full control over the information content presented to participants is therefore necessary to credibly identify these effects.

The survey begins by eliciting participants' beliefs about the size of the overall gender wage gap. To establish a common baseline, all participants are subsequently provided with the same empirical estimate, for which the percentage explained by children is later calculated. Participants are then introduced to the concept of the child-related component of the wage gap and to how the share explained by children is calculated. In the first stage of randomization, participants are assigned either to a *High estimate: 90* condition, in which children are assumed to affect both women's and men's wages, yielding an estimate of 90 percent as calculated in Cortés and Pan (2023) or to a *Low estimate: 60* condition, in which children are assumed to affect only women's wages, yielding an estimate of 60 percent. Both estimates are calculated using the same sample from the Panel Study of Income Dynamics (PSID) and the same underlying estimate of the overall gender wage gap. The difference between them arises solely from whether the effect of children on men's wages is included when computing the child-related component. After reporting their beliefs about the share of the gap attributable to children, participants are randomized a second time into either a *Control* group, which is reminded of their own prior belief, or an *Info* group, which is shown the corresponding empirical estimate.

This two-stage randomization creates exogenous variation in beliefs about the contribution of children to the gender wage gap between the *Info* & *60* and *Info* & *90* groups, as the only difference between these groups is the magnitude of the share they are told is explained by children. Comparing these groups therefore identifies the causal effect of beliefs about the importance of children for policy preferences. Comparing the *Info* groups to the *Control* group instead allows us to isolate the effect of receiving information that differs from one's prior belief and thereby shifts beliefs upward or downward, relative to the effect of merely being made aware of the potential role of children in explaining the gap.

Following the intervention, we elicit participants' overall support for reducing the gender wage gap and their preferences regarding whether pay differences related to children should be prioritized over pay differences related to other factors. We then measure support for six specific policies that are central to policy debates and that map directly into alternative explanations of the wage gap. Child-related policies include child care subsidies, paid maternity leave, and paid paternity leave, which aim to reduce career penalties associated with parenthood and to reallocate care responsibilities. Discrimination-related policies include gender quotas, affirmative action plans, and equal pay legislation, which target unequal treatment in the labor market independent of children. Policy preferences are measured on a 5-point Likert scale. In addition, participants are asked whether they would be willing to donate to Child Care Aware of America (CCAoA).

Finally, we elicit several additional perceptions and beliefs that may shape policy preferences or mediate the relationship between beliefs and these preferences. Most importantly, we measure how fair and efficient participants perceive the child-related part of the gender wage gap to be compared to the part unrelated to children.

We begin by documenting that beliefs about the share of the gender wage gap attributable to children vary widely in our sample. On average, individuals estimate this share to be around 50%. This perceived value is substantially lower than the empirical estimate of 90% reported in Cortés and Pan (2023) and lower than the corresponding estimate of 60% obtained when abstracting from the small positive effect of children on men's earnings.

Turning to the effects of the information intervention, we causally identify how beliefs about the extent to which children contribute to the gender wage gap affect policy preferences by comparing the treatment effects of the *Info & 60* and *Info & 90* conditions. This analysis shows that a higher perceived contribution of children to the gender wage gap significantly increases individuals' preference to prioritize child-related pay differences and reallocates support across policy types, shifting it toward child-related policies and away from discrimination-related ones. This pattern is primarily driven by higher support for maternity leave and lower support for quotas and affirmative action programs.

Moreover, by grouping participants in the *Info* groups based on whether the empirical estimate shown to them is above or below their stated prior belief, and comparing them to individuals with similar priors in the *Control* group, we investigate how the direction of the information signal shapes treatment effects. This comparison allows us to isolate the effect of receiving information that differs from one's prior beliefs, relative to the effect of merely being made aware of the potential role of children in explaining the gap. Using this variation, we show that treatment effects, especially on support for discrimination-related policies and relative support, depend significantly on the direction of the information signal relative to individuals' prior beliefs. When participants receive information below their prior, relative support shifts away from child-related policies and toward discrimination-related policies. When the information implies a larger role of children, individuals instead increase their preference for prioritizing child-related pay differences.

Finally, we examine whether the impact of information on policy support depends on individuals' fairness perceptions. By shifting beliefs about the contribution of children to the gender wage gap, we alter the perceived importance of child-related factors relative to other explanations. This, in turn, may influence policy support through individuals' perceptions about the fairness of child-related pay differences compared to other explanatory factors. We test this by comparing treatment responses between participants who perceive the child-related part of the gender wage gap as less fair than the part not explained by children and those who view it as equally or more fair. This analysis shows that treatment effects, particularly on overall support, depend significantly on fairness perceptions, with individuals who perceive the child-related component as less fair exhibiting a more positive relationship between overall support and the direction of the information signal. In particular, general support increases when individuals are informed that the component they consider less fair accounts for a larger share of the wage gap. By contrast, when they are led to believe that the component they perceive as equally or more fair is larger, general support remains unchanged.

To the best of our knowledge, this is the first paper to experimentally shift beliefs about the causes of the gender wage gap, in particular the role of children in driving it, and to examine how these shifts affect the overall level of support for policies addressing gender inequality and the relative support for different types of policies. By doing so, our paper contributes to several strands of literature.

First, we build on work documenting the central role of children in driving gender inequality. A substantial body of empirical research documents that children have a large impact on women's career trajectories but not on men's (see, e.g., Cortés and Pan, 2023; Goldin, 2024; Olivetti, Pan, and Petrongolo, 2024, for reviews). More recent work demonstrates that this differential impact accounts for much of the remaining gender earnings gap in high-income countries (e.g., Kleven, Landais, and Søgaard, 2019; Cortés and Pan, 2023; Kleven, Landais, and Leite-Mariante, 2025). Women continue to be the primary caregivers. As a result, both demand-side factors, such as discrimination based on marital status and parenthood (Correll, Benard, and Paik, 2007; Becker, Fernandes, and Weichselbaumer, 2019), and supply-side factors, including career interruptions, reduced working hours, and sorting into more flexible but lower-paying jobs (Bertrand, Goldin, and Katz, 2010; Goldin, 2014; Hotz, Johansson, and Karimi, 2018; Kleven, Landais, and Søgaard, 2019), have been shown to be important mechanisms underlying the differential impact of children on women's and men's careers.

While previous research has examined the mechanisms behind the child penalty and its role in explaining the gender earnings gap, less is known about how the public perceives its importance. In particular, we show that people substantially underestimate how much of the gender wage gap is driven by children, with a larger misperception observed among those without children. This finding is consistent with evidence from Kuziemko et al. (2018), who show that women underestimate the effect of motherhood on their employment. More broadly, our paper connects to a growing literature examining expectations about labor market consequences of motherhood (Kuziemko et al., 2018; Gong, Stinebrickner, and Stinebrickner, 2022; Caplin, Leth-Petersen, and Tonetti, 2025) and the perceived pecuniary and non-pecuniary implications of mothers' labor supply choices (Boinet et al., 2024; Costa-Ramón et al., 2024; Boneva et al., 2026). We contribute to this literature by studying beliefs about the extent to which parenthood explains the gender wage gap and by examining how these beliefs shape policy preferences.

In doing so, our work also relates to a broader literature that identifies how beliefs about inequality shape related policy preferences (e.g., Kuziemko et al., 2015; Alesina, Stantcheva, and Teso, 2018; Alesina, Ferroni, and Stantcheva, 2021; Haaland and Roth, 2023; Günther and Martorano, 2025; Henkel et al., 2025). Within this stream, our paper is most closely related to Settele (2022) and Casarico, Schuetz, and Uebelmesser (2024), who also examine how beliefs about gender inequality influence support for policies aimed at reducing it. Both studies focus on how beliefs about the size of inequality affect policy preferences, with Settele (2022) analyzing beliefs about the size of the gender wage gap and Casarico, Schuetz, and Uebelmesser (2024) considering beliefs about both the size of the gender earnings gap and the size of the gender pension gap. In contrast, we examine beliefs about the causes of gender inequality, focusing on the perceived importance of children relative to other factors. We study how shifting beliefs about its causes influences both overall support for policies addressing gender inequality and support for specific policies.

We are not aware of any other studies that fix the size of an inequality and shift beliefs about the percentage explained by a particular factor. Furthermore, we introduce the concept of relative support, capturing how individuals allocate support between policies targeting the child-related component of the gender wage gap and those targeting other contributing factors.

Lastly, we zoom in on perceptions of fairness and efficiency that mediate how beliefs about the contribution of children to the gender wage gap translate into policy preferences. Our results connect to a broader literature showing that fairness views are shaped by the perceived causes of inequality and that these views, in turn, affect the acceptance of inequality (see, e.g., Almås, Hufe, and Weishaar, 2025, for a review). We contribute to this literature by connecting beliefs about the importance of an observable determinant of inequality, the contribution of children to the gender wage gap, to perceptions of its fairness relative to other factors driving inequality, and by demonstrating how these two dimensions jointly influence policy preferences.

The remainder of the paper proceeds as follows. Section 1.2 outlines the experimental design and discusses the treatments and elicited variables in detail. Section 1.3 describes the sample and provides descriptive evidence on prior beliefs regarding both the size of the gender wage gap and the percentage explained by children. Section 1.4 presents the results, and Section 1.5 concludes.

1.2 Experimental design

Studying individuals' beliefs about the percentage of the gender wage gap explained by having children and how these beliefs relate to policy views is challenging using observational data alone. Pre-existing beliefs and selective exposure to information complicate the identification of causal effects. Hence, an experimental setting that allows for exogenous variation in individuals' beliefs, precise measurement of political preferences and underlying perceptions, and full control over the information content presented to participants is necessary. Specifically, an ideal experimental design should: (i) establish a common baseline by anchoring participants' beliefs about the overall size of the gender wage gap, ensuring that subsequent information about shares is interpreted consistently across individuals; (ii) clearly explain the concept of and elicit prior beliefs about the percentage of the wage gap explained by having children and ensure understanding; (iii) randomly assign participants to receive an empirically obtained estimate of this percentage and randomly vary the magnitude of the estimate presented, thereby generating exogenous variation in beliefs about the child-related component of the wage gap, and assess belief updating in response; (iv) include a comprehensive set of survey items to elicit political preferences and measure underlying beliefs and perceptions through which information may affect policy preferences.

1.2.1 Overview and timeline

We conducted an online survey experiment using the U.S. participant pool of the platform Prolific. In the experiment, participants are asked a series of questions regarding their beliefs and preferences. Figure 1.1 presents the experimental setup.

At the beginning of the survey, subjects have to pass a basic attention and comprehension check to continue with the main part. The main part starts with participants being asked to give their best guess about the size of the gender wage gap. They are then informed about the size, calculated using the Panel of Income Dynamics (PSID)[1], and subsequently asked to report their narratives about what causes the gender wage gap in an open-text question format.

Participants are then randomly assigned to one of two groups: *Low estimate: 60* or *High estimate: 90*. These two estimates result from a slight variation in how the child-related part of the gender wage gap is defined—specifically, whether or not the effect of children on men's wages is included. Both groups receive an introduction to the child-related part of the gender wage gap, including an explanation of how the percentage of the gap explained by children is calculated. This information is tailored to the estimate assigned to each group and followed by a comprehension question. The only difference between the groups lies in the information provided on this introduction page. Section 1.2.2 provides further details on these treatment groups. Afterward, all participants are asked to provide their own estimate of the percentage of the gender wage gap explained by children.

Participants in each of the two groups, *Low estimate: 60* and *High estimate: 90*, are further divided into a *Control* group and an *Info* group, resulting in four treatment conditions: *Info* & *60*, *Info* & *90*, *Control* & *60* and *Control* & *90*. The *Info* groups receive information about the empirically estimated share of the gender wage gap explained by having children, corresponding to the estimate of the treatment group to which they were assigned. Participants in the *Control* groups do not receive any new information. Instead, they are reminded of their own estimates. We then check whether participants have understood the treatment and paid attention. Afterwards, they answer questions regarding their policy views, which include our main outcome variables and are discussed in more detail in Section 1.2.3. We then elicit several additional underlying perceptions and beliefs that may shape policy preferences or mediate the relationship between beliefs and these preferences. This includes their

1. The estimate is based on the cleaned data used by Cortés and Pan (2023) from the PSID for the years 2005 to 2016.

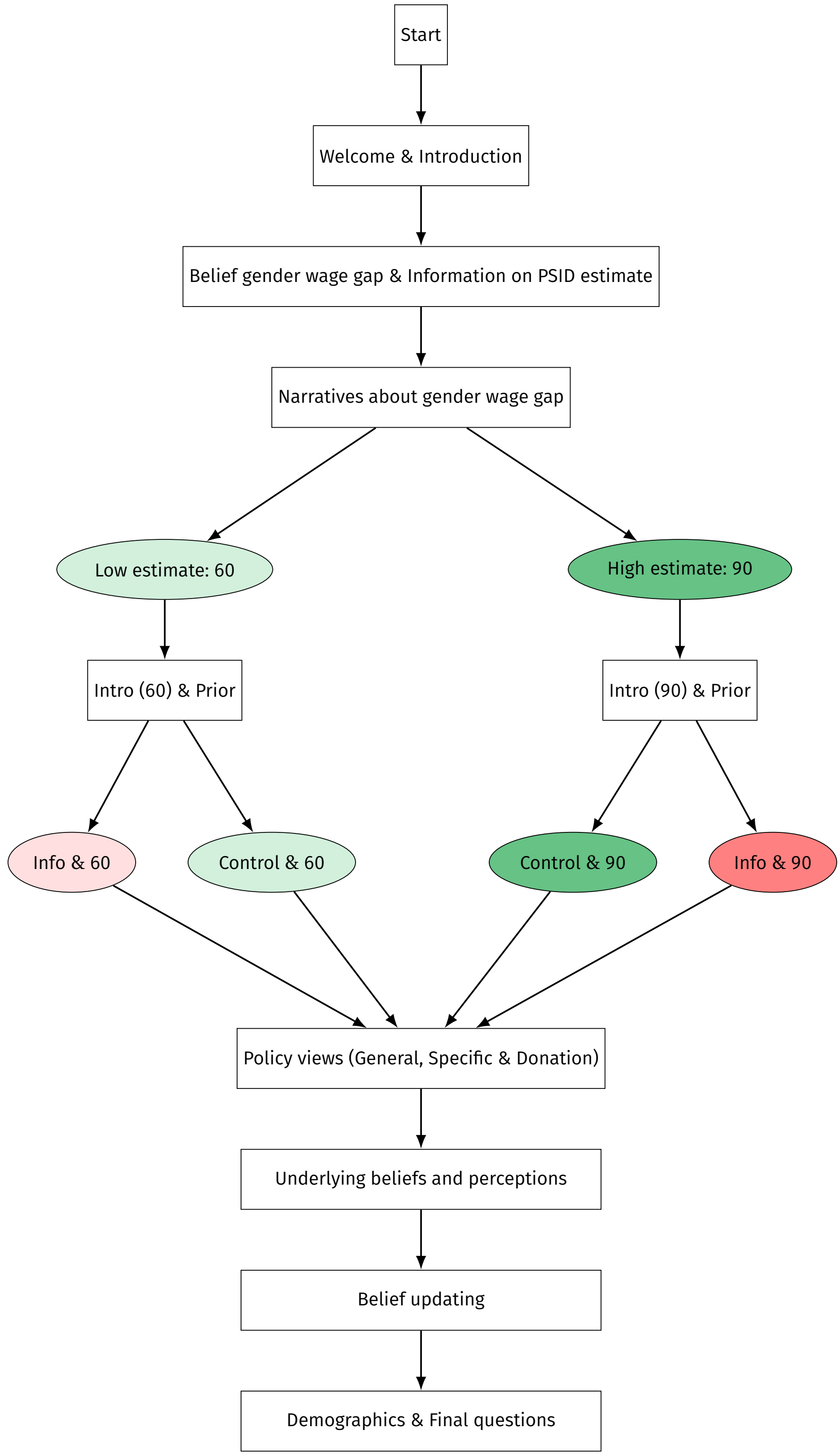


Figure 1.1. Experimental timeline

fairness and efficiency perceptions, both regarding the overall gender wage gap and the child-related component relative to the part unrelated to children, as well as their views on the sources of inequalities underlying the gender wage gap and the extent to which these sources are attributed to biological or societal influences. Next, we elicit their beliefs about the percentage of the gap explained by children again to assess belief updating. Finally, participants provide demographic information and finish the survey with a few final questions.

1.2.2 Treatment conditions

We implement four treatment conditions that vary along two key dimensions: (i) the assumptions underlying the calculation of the percentage of the gender wage gap explained by having children, which result in two different empirical estimates (*Low estimate: 60* vs. *High estimate: 90*), and (ii) whether participants are provided with this empirical estimate (*Control* vs. *Info*). Participants are first randomly assigned to either the *Low estimate: 60* or the *High estimate: 90* condition with equal probability. Within each estimate condition, one-third of participants are assigned to the *Control* group and two-thirds to the *Info* group. This results in one-third of the total sample being allocated to each of the *Info* & *60* and *Info* & *90* conditions, and one-sixth to each of the *Control* & *60* and *Control* & *90* conditions. These four treatment conditions allow us to independently vary (i) whether participants receive information about the empirical estimate and (ii) the size of that estimate. This two-stage randomization generates exogenous variation in beliefs about the contribution of children to the gender wage gap between the *Info* & *60* and *Info* & *90* groups. We therefore compare these groups to identify the causal effect of beliefs about the importance of children on policy preferences.

In addition, comparing individuals in the *Info* groups with those in the *Control* groups who hold similar prior beliefs allows us to isolate the effect of receiving information that shifts beliefs relative to one's prior belief, rather than the effect of merely being made aware of the potential role of children in explaining the gender wage gap.

1.2.2.1 *Low estimate: 60* vs. *High estimate: 90*

The two different estimates result from a slight variation in how the child-related part of the gender wage gap is defined – specifically, whether or not the effect of children on men's wages is included. The high estimate corresponds to the one reported in Cortés and Pan (2023). To obtain a second estimate, we redefined the child-related part to reflect only the impact of children on women's wages, rather than the differential impact on women's and men's wages. Since men's wages tend to slightly increase after having children, including this effect leads to a higher estimate of the percentage of the gender wage gap explained by children (90), while excluding it yields a lower estimate (60) (see Appendix 1.A.2 for further details).

Participants in all conditions read and must understand a brief explanation of how the percentage is calculated before reporting their belief. To make the two treatments as comparable as possible, we defined the percentage explained by children in exactly the same way, but instructed participants in *Low estimate: 60* to assume that men's wages are unaffected by children, so they would consider only the effect of children on women's wages when estimating the percentage. Participants in *High estimate: 90* were instead instructed to assume that men's wages slightly increase after having children (see Figure 1.A.1 for the exact wording).

1.2.2.2 ***Control* vs. *Info***

The *Control* and *Info* treatments differ only in whether participants are provided with the empirical estimate calculated using the PSID data (see Figure 1.A.2). The size of the estimate shown depends on whether participants are in the *Low estimate: 60* or the *High estimate: 90* condition.

1.2.3 Variables

The questions asked in the survey can be categorized into the following groups of variables: Prior beliefs, posterior beliefs, policy preferences, perceptions about the child-related part of the gender wage gap, beliefs and perceptions about the overall gender wage gap, understanding of information and additional variables. Table 1.A.1 and Table 1.A.2 present the variables used and outline their exact question and response format. Screenshots of the entire experiment can be seen in Appendix 1.A.4.

1.2.3.1 Prior beliefs

The first set of variables captures participants' prior beliefs about the gender wage gap and its underlying causes. To establish a common baseline, we elicit participants' beliefs about the size of the gender wage gap by asking: *"How many dollars do you think a woman earns for every $100 earned by a man?"*. Responses are given on a scale ranging from $0 to $200. We then provide participants with an anchor by informing them of the empirical estimate used in Cortés and Pan (2023) to calculate the percentage of the gender wage gap explained by children.

The prior belief about the percentage of the gap explained by children is one of the key explanatory variables in our analysis, from here on simply referred to as prior. It is used to examine its correlation with policy preferences in the pure *Control* group and to determine the direction of belief updating for participants who receive information. Participants are asked *"What percentage of the gender wage gap do you think is explained by children?"* and their answers are bound by a slider between 0% and 100%.

Participants are randomly assigned to be incentivized either for their belief about the size of the gender wage gap or for their belief about the percentage of the gap explained by having children. Incentives are provided in the form of a $1 bonus for answers within ±5 percentage points of the estimate based on the PSID data.

In addition, participants answer an open-ended question eliciting their beliefs about the causes of the gender wage gap. Specifically, we ask: *"Why do you think women earn on average less than men?"*.

1.2.3.2 Posterior beliefs

At the end of the survey, participants are again asked to report their own belief about the percentage of the wage gap explained by children, from here on referred to as posterior. This posterior allows us to assess whether participants accepted the information and updated their belief accordingly. In addition, we elicit whether participants believe that children are an important explanation for the gender wage gap, reflecting the subjective importance they place on this factor.

1.2.3.3 Policy preferences

This set of variables captures participants' policy preferences, which serve as the main outcome variables. The outcomes include overall policy support, specific policy support and relative policy support.

Overall support is reflected by general support for reducing the gender wage gap and support for government intervention to address the gap. More specifically, participants answer the questions *"How important do you believe it is for society to work toward reducing the gender wage gap?"* and *"To what extent do you support government action to reduce the wage gap?"* on a 5-point Likert scale.

To address specific support, we ask participants to indicate the level of support for several specific policies, which is again measured on a 5-point Likert scale. These policies can be grouped into child-related policies (maternity leave, paternity leave, and child care subsidies) and discrimination-related policies (equal pay legislation, affirmative action, and gender quotas). We also elicit support for other policies, namely: abortion rights and unrelated policies, such as: infrastructure investment, defense, and healthcare spending. In addition, participants are asked whether they would be willing to donate to a national organization that supports all aspects of the child care system - Child Care Aware of America (CCAoA). They are informed that one randomly selected participant would win $100. Participants are then asked to decide how much of the prize they would like to keep and how much they would like to donate to the CCAoA if they were to win. To encourage donations, we subsidized the donation by adding 50% to the amount participants chose to give.

An important feature of our design is that we directly elicit participants' relative support for child-related policies compared to other policies by asking the question *"Do you think the government should prioritize reducing gender pay differences caused by children or those caused by other factors?"* on a 5-point Likert scale. Moreover, we construct their relative support for child-related policies compared to discrimination-related policies by calculating the difference between their average support for child-related policies and their average support for discrimination-related policies.

1.2.3.4 Perceptions about the child-related part

Perceptions and opinions about children as an explanation for the gender wage gap may vary even among participants with similar beliefs about the share of the gap explained by children. The relationship between such beliefs and policy preferences may depend on these underlying perceptions. To capture this, we ask participants to assess how fair and efficient[2] they consider the child-related component of the gender wage gap to be, compared to the portion not explained by having children. This gives us the notion of relative fairness and relative efficiency.

1.2.3.5 Beliefs and perceptions about the overall gender wage gap

This set of variables captures beliefs and perceptions about the overall gender wage gap that are elicited after the treatment. The information provided may influence policy preferences through these channels. More specifically, we elicit participants' perceived fairness and efficiency of the overall gender wage gap, as well as their views on its underlying sources.

1.2.3.6 Understanding of information

After either being reminded of their initial guess (*Control* group) or receiving information about the percentage of the gender wage gap explained by having children (*Info* group), participants are asked to report both the percentage of the gender wage gap explained by children and the percentage not explained by children. For participants in the *Control* group, this serves to check whether they understand that these two percentages must sum to 100. For participants in the *Info* groups, this additionally verifies whether they paid attention to the information provided.

1.2.3.7 Additional variables

We collect additional variables that cannot be influenced by our treatment, but may be significantly correlated with policy preferences. They include demographics, trust in government, gender norms, and zero-sum thinking.

2. Efficiency here refers to the optimal allocation of talent, that is, whether pay differences arise because men and women are inherently better suited to different types of work.

1.3 Descriptive evidence

1.3.1 Sample description

The experiment was conducted in May 2025 using the U.S. sample of the platform Prolific. The survey was completed by 1,498 participants.[3] After excluding observations as outlined in the pre-registration, that is, subjects who are likely to be bots, who fail a comprehension check more than twice or fail an attention check midway through the survey, we are left with a final sample size of 1,274.[4] After applying the exclusion criteria, the randomization into the four different treatment groups is still balanced with 33% of the participants being allocated to the *Control* group (*Control* & *90*: 17% and *Control* & *60*: 16%), 32% being allocated to the *Info* & *90* group and the remaining 35% to the *Info* & *60* group.

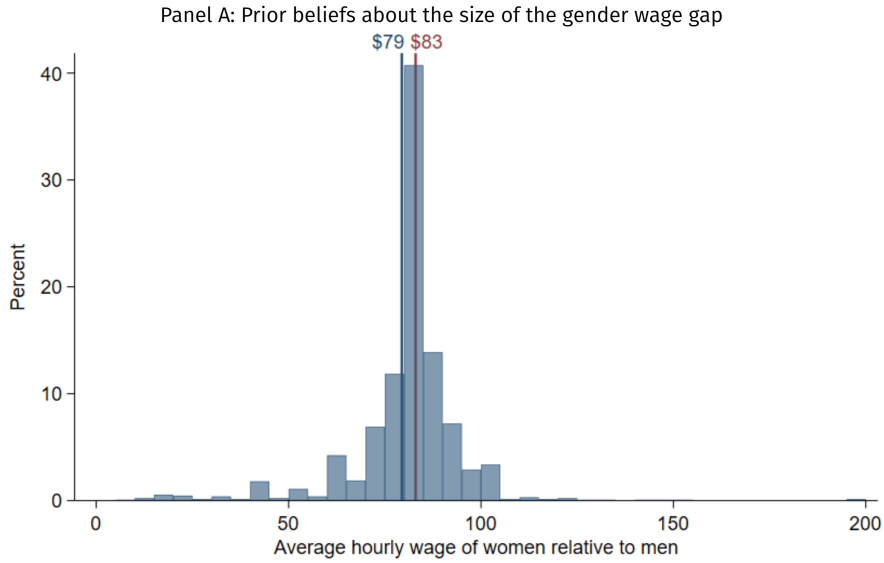
Table 1.B.1 presents the demographic characteristics by treatment group used in the analysis, namely the *Control*, *Info* & *60*, and *Info* & *90* groups. Overall, demographic characteristics are largely balanced across treatment groups. The main differences are that the *Info* & *60* group contains slightly more married participants, a difference that approaches statistical significance when compared to the *Info* & *90* group ($p = 0.072$). The *Control* group includes slightly more unemployed individuals, a difference that is statistically significant when compared to the *Info* & *60* group ($p = 0.049$). In addition, the *Control* group has slightly fewer female and slightly more male participants compared to the *Info* groups; however, pairwise comparisons indicate that these differences are not statistically significant.

1.3.2 Prior beliefs

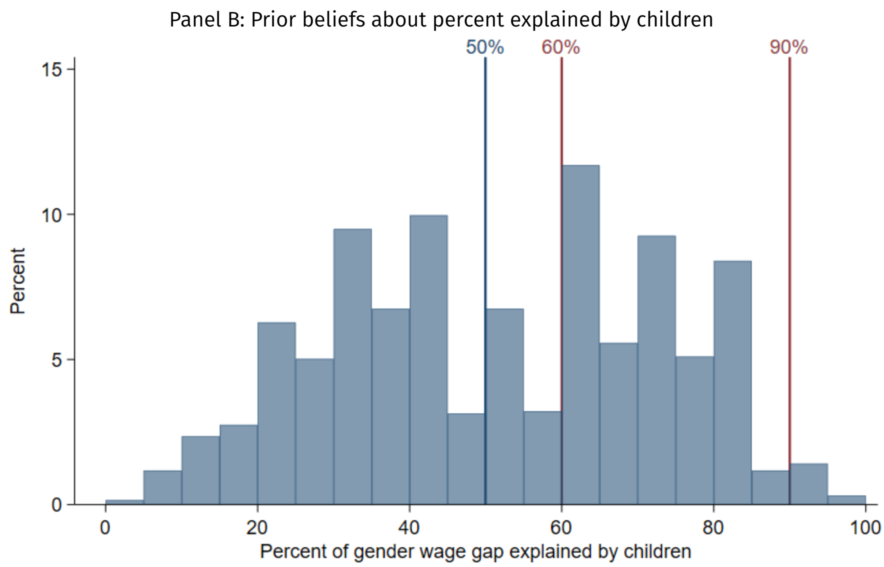
We now turn to examining how well participants are informed about both the size of the gender wage gap and the percentage of the gap explained by having children. Panel A in Figure 1.2 shows the distribution of participants' beliefs about the size of the gender wage gap, expressed as women's average hourly wage relative to men's. The distribution is centered around a mean of 79, meaning that women earn $79 for every $100 earned by a man. This guess is close to the empirically estimated value of $83 based on the PSID. Notably, the standard deviation of the distribution is $15. This suggests that the public is generally well informed about the size of the gender wage gap. After reporting their guess, we anchor their belief about the size of the gender wage gap by providing all participants with the empirical estimate obtained by using the PSID.

3. We pre-registered a target sample size of 1,500 observations. Ex post, we found that 2 participants, although recorded on Prolific, were not present in our dataset.

4. In addition to the pre-registered exclusion criteria, we excluded 9 participants who participated twice after having failed the initial comprehension or attention check on their first attempt.



Notes: This figure shows the distribution of responses to the question of how many dollars respondents think a woman earns for every $100 earned by a man. Responses were provided using a slider ranging from $0 to $200. The blue line marks the mean of the distribution ($79). The red line marks the value of the empirical estimate provided ($83).



Notes: This figure shows the distribution of responses to the question of what percentage of the gender wage gap respondents believe is explained by children. Responses were provided using a slider ranging from 0% to 100%. The blue line marks the median and mean of the distribution (50%). The red line marks the values of the low (60%) and high (90%) estimates provided to the *Info* groups.

Figure 1.2. Distribution of prior beliefs

While participants are well-informed about the gender wage gap, their beliefs about the share explained by having children vary widely. Panel B in Figure 1.2 shows that prior beliefs are widely dispersed with a mean (and median) of 50 and a standard deviation of 21 percentage points. This dispersion is not driven by the differential assumptions in the *Low estimate: 60* and *High estimate: 90* conditions. Beliefs are similarly distributed across both groups, with means of 49 and 51 and standard deviations of 22 and 21, respectively (see Figure 1.B.1 for the corresponding distributions across groups).[5] Importantly, on average participants' priors are significantly lower than the empirically estimated values derived from the PSID ($p <$ 0.0001). The same pattern emerges when comparing priors within each condition to the corresponding estimate, with both differences highly significant ($p < 0.0001$). While 98% of participants report a value below 90, 57% report a value below 60, implying that 43% report a value greater than or equal to 60. This indicates substantial scope for informing participants and shifting beliefs in both directions.

The dispersion in priors does not appear to reflect misunderstanding of the concept. Instead, it likely reflects heterogeneous beliefs about the role of children in explaining the gender wage gap. To ensure that participants in the final sample understood the core concept, we applied a conservative exclusion criterion based on a comprehension check regarding the percentage explained by children. After entering their prior belief and either being reminded of it or receiving new information, participants are again asked to recall what percentage is explained by children and what percentage is not. Of those remaining in the final sample, 94% passed the understanding checks—defined as providing two percentages that sum to 100 and correctly recalling either their own prior estimate or the PSID estimate presented to the *Info* groups. Therefore, it seems likely that the participants in our experiment were just not aware of this percentage prior to treatment.

Supporting the notion that participants are not well-informed about the role children play in the gender wage gap, only 28.34% referred to children or, more broadly, family responsibilities in the open-text question asking why they believe women earn less on average. Among these respondents, 24.88% described direct effects, such as women taking career breaks or reducing hours, while 5.34% referred to indirect effects, such as employers discriminating against women because of their expected roles as primary caregivers. Respondents who mentioned children estimate the share of the wage gap explained by children to be 5.14 percentage points higher than those who did not ($p = 0.0001$). This pattern further suggests that the elicited prior captures beliefs about the importance of children in explaining the gender wage gap.

5. The difference in means is weakly significant ($p = 0.089$), but becomes statistically insignificant once controlling for demographic characteristics.

1.3.2.1 Priors by demographics

We now turn to the question of whether some groups are better informed than others. While prior beliefs about the percentage of the gender wage gap explained by having children are significantly correlated with gender, political affiliation, and ethnicity, parental status emerges as the strongest predictor (see Table 1.B.2). People with at least one child tend to report significantly higher prior estimates than their childless counterparts. The tendency of childless individuals to underestimate the impact of children on women's wages is consistent with Kuziemko et al. (2018), who document that women without children underestimate the employment effects of motherhood.

1.3.2.2 Priors by treatment group

To generate two distinct estimates, we varied whether the effect of children on men's wages was included when calculating the percentage of the gender wage gap explained by children. The only difference between the *Control* & *60* and *Control* & *90* groups lies in the assumption about men's wages they were told to make. If this assumption did not influence participants' prior beliefs, we can pool these into a single *Control* group that receives no additional information. To assess this, we test whether prior beliefs differ across the two groups. Table 1.B.3 reports mean prior beliefs for the two *Control* groups. Tests of equality of means show no statistically significant differences, indicating that the assumptions did not meaningfully affect priors and supporting the pooling of the two groups into one *Control* group.

1.4 Results

In this section, we examine whether and how beliefs about the importance of children as a driver of the gender wage gap influence policy views. The section is structured as follows. First, we assess whether the information treatments successfully shifted participants' beliefs about the size of the child-related part of the gender wage gap.

Next, we analyze whether the perceived importance of children affects overall policy support, support for specific policies, and support for child-related policies relative to other types of policies. We do so by examining the effects of the information provided to respondents, which varied in the extent to which the gender wage gap was attributed to children. We then extend the analysis by considering the relationship between participants' prior beliefs and the information they received.

Finally, we explore how perceptions of fairness influence these relationships. Specifically, we investigate whether and how the perceived fairness of the child-related part of the gender wage gap relative to the perceived fairness of other contributing factors mediates the link between the belief about the importance of children and policy views.

1.4.1 Updating by treatment

To examine whether and how beliefs about the importance of children as a driver of the gender wage gap influence policy views, we compare participants across three treatment conditions: the *Control* group, which did not receive any information about an empirical estimate of the size of the child-related part of the gender wage gap, and two information treatments, *Info* & *60* and *Info* & *90*, which received empirical estimates of 60% and 90%, respectively. Before turning to participants' policy views, we first assess whether the groups differed in their prior beliefs and whether the treatment successfully shifted beliefs.

Table 1.1. Updating by treatment groups

	Treatment group		
	Control	Info & 60	Info & 90
Prior	49.402	49.261	51.312
	(21.997)	(21.663)	(20.480)
Posterior	50.043	59.388	84.788
	(22.160)	(8.217)	(15.787)
Observations	423	441	410

Notes: This table reports mean prior and posterior beliefs about the share of the gender wage gap explained by children for the treatment groups used for the main analysis.

Table 1.1 reports mean prior and posterior beliefs by treatment status. Tests of equality of means show no statistically significant differences in prior beliefs. In particular, priors do not differ between the pooled *Control* group and the two *Info* groups ($p = 0.924$ and $p = 0.195$, respectively), nor between *Info* & *60* and *Info* & *90* ($p = 0.157$). Given the absence of meaningful differences in priors, we proceed under the assumption that any subsequent differences between the *Control*, *Info* & *60*, and *Info* & *90* groups are driven by the information participants received.

Next, we examine whether the treatment successfully shifted beliefs. The results in Table 1.1 show clear differences in posterior beliefs across treatment conditions. Posterior beliefs in the *Info* groups closely correspond to the information they received, with mean posterior beliefs of 59.39 in *Info* & *60* and 84.79 in *Info* & *90*. In contrast, mean posterior beliefs in the *Control* group remain at 50.04 and do not significantly differ from priors.

1.4.2 Beliefs about importance of children and policy support

Having established that the treatment successfully shifts participants' beliefs about the importance of children in explaining the gender wage gap, we now examine whether and how this perception influences individuals' policy preferences. We begin by analyzing how the information provided to respondents affects various measures capturing their policy views. We do this by estimating the following specification:

$$Y_i \;=\; \beta_0 + \beta_1 \cdot \mathit{Prior}_i + \beta_2 \cdot \mathit{Info}_i + \beta_3 \cdot \mathit{Info}\ \&\ 90_i + \beta_4 \cdot \mathbf{X}_i + \epsilon_i \tag{1.1}$$

where Y denotes the outcome variable capturing the respondent's policy view, *Prior* represents the respondent's initial belief about the share of the gender wage gap explained by children, *Info* is a dummy variable indicating whether the respondent was provided with an empirical estimate of this share, *Info* & *90* is a dummy variable equal to one if the respondent was randomized into the *Info* & *90* treatment group, and **X** is a vector of demographic controls.

The coefficient β_2 measures the difference between the *Control* group and the *Info* & *60* group. This coefficient captures a combination of effects. It reflects the overall impact of receiving information, which may not only shift beliefs but also reduce uncertainty about the importance of children in explaining the gender wage gap. It also encompasses potentially heterogeneous effects, as both the extent and direction of belief updating depend on respondents' priors. The coefficient β_3, by contrast, measures the difference between the *Info* & *60* and *Info* & *90* groups. The only distinction between these groups is the magnitude of the share they are told is explained by children. Regardless of their priors, individuals in the *Info* & *90* group are informed that a larger share of the gender wage gap is attributable to children. Comparing these two groups therefore allows us to identify the causal effect of beliefs about the extent to which children contribute to the gender wage gap on policy preferences. The sum of the two coefficients $\beta_2 + \beta_3$ therefore measures the difference between the *Control* group and the *Info* & *90* group.[6]

Table 1.2 presents results for several outcome variables capturing the different aspects of policy preferences. Columns (1)–(2) report results for overall support for reducing the gender wage gap, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)). Columns (3)–(4) then examine how this overall support translates into support for specific policies. Column (3) reports results for child-related policies, measured as the average support for maternity leave, paternity leave, and child care subsidies, while column (4) reports results for discrimination-related policies, measured as the average support for affirmative action plans (AAP), gender quotas, and equal pay legislation. Rather than shifting overall support, the information treatments may shift support across policies targeting different aspects of the gender wage gap, with increases in support for some offsetting decreases for others. Columns (5)–(6) therefore report results for relative measures of policy support. Column (5) shows the difference in support between child-related and discrimination-

6. When reporting differences between the *Control* group and the *Info* & *90* group, we perform a Wald test of the null hypothesis that $\beta_2 + \beta_3 = 0$.

Table 1.2. Policy support by treatment status

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Prior	-0.002	-0.002	-0.001	-0.001	-0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Info	0.150**	0.056	-0.103*	0.045	-0.148***	0.132*
	(0.061)	(0.061)	(0.060)	(0.051)	(0.053)	(0.071)
Info & 90	-0.027	-0.069	0.056	-0.105**	0.161***	0.227***
	(0.059)	(0.062)	(0.068)	(0.051)	(0.058)	(0.079)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.4. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children.[7]

Starting with the comparison between *Info* & *60* and *Info* & *90*, two key findings emerge regarding how beliefs about the role of children in explaining the gender wage gap shape policy preferences. First, on average, overall support is not significantly affected by the perceived importance of children as an explanation for the gender wage gap. If anything, the relationship appears slightly negative. Second, the perceived importance of children as an explanatory factor significantly influences the relative support for different types of policies. The results in Column (6) indicate that a higher perceived importance of children leads to a significant increase in the preference for prioritizing pay differences related to having children, with the *Info* & *90* group exhibiting a 0.227 s.d. higher preference relative to *Info* & *60* ($p < 0.01$). Similarly, Column (5) shows relatively greater support for child-related

7. The last two outcomes were not explicitly pre-registered. However, relative support was directly measured in the survey and closely corresponds to the pre-registered measures of support for specific policies. The observed pattern is consistent with the effects across those individual items.

policies compared to discrimination-related ones, with the *Info & 90* group showing a 0.161 s.d. higher level of relative support ($p < 0.01$). This pattern is primarily driven by a significantly negative relationship between the perceived importance of children and support for discrimination-related policies (Column (4)), where the *Info & 90* group reports a 0.105 s.d. lower level of support relative to the *Info & 60* group ($p < 0.05$).

To examine which specific policies are most affected, Table 1.B.4 reports results for each policy separately. For child-related policies, participants in the *Info & 90* group show significantly greater support for maternity leave (0.148 s.d.; $p < 0.10$), but not for paternity leave or child care subsidies. Similar to the support for child care subsidies, donations to the overall child care system do not differ significantly between the groups (see Table 1.B.5). For discrimination-related policies, participants in *Info & 90* exhibit significantly lower support for quotas (−0.102 s.d.; $p < 0.10$), and especially for AAP (−0.118 s.d.; $p < 0.05$), while support for stricter equal pay legislation is only marginally lower (−0.093; $p = 0.137$).

Result 1 Higher perceived contribution of children to the gender wage gap significantly reallocates support across policy types, shifting it toward child-related policies and away from discrimination-related ones. This pattern is primarily driven by higher support for maternity leave and lower support for quotas and affirmative action programs.
However, beliefs about the extent to which children contribute to the gender wage gap do not, on average, affect overall support for reducing the gender wage gap.

We next examine whether the main results are driven by specific subgroups or instead reflect broadly similar responses across demographic characteristics. Across heterogeneity analyses, we find no significant differences between demographic groups in the shift in relative support from discrimination-related policies toward child-related policies in response to a higher perceived importance of children as an explanation for the gender wage gap, though some differences emerge for other policy outcomes. Interacting the treatment with gender shows that the only significant difference in the relationship between beliefs and policy views arises for the preference to prioritize pay differences related to having children, where the positive response is primarily driven by men (see Table 1.B.6 and Table 1.B.7). Results by parental status show largely similar responses across groups, though parents tend to exhibit a somewhat more positive relationship between beliefs about the role of children and both overall support and support for specific policies, which only reaches statistical significance for child care subsidies (see Table 1.B.8 and Table 1.B.9). Finally, heterogeneity by party affiliation indicates largely similar relationships between beliefs and relative policy support across groups. Differences arise only for overall support for government intervention, where individuals aligned with the Democratic Party respond more positively to a higher perceived importance of children (see Table 1.B.10 and Table 1.B.11).

We now turn to a discussion of one potential channel that may contribute to the treatment effect. Individuals who hold similar beliefs about the extent to which children explain the gender wage gap may differ in how they attribute this effect to penalties in women's wages versus gains in men's wages. Part of the experimental variation that creates exogenous differences in beliefs between the *Info* & *60* and *Info* & *90* groups stems from the different assumptions participants were asked to make about men's wages. Differences in policy preferences across these groups may therefore partly reflect differences in attribution. However, we cannot directly isolate this channel when comparing the *Info* & *60* and *Info* & *90* groups, as these treatments simultaneously shift both the perceived size of the child-related component and the underlying assumptions about men's wages.

To provide suggestive evidence on whether attribution affects policy views, we compare the two *Control* groups. As both groups hold similar priors about the overall share explained by children, the only difference between *Control* & *60* and *Control* & *90* is the assumption about men's wages. Any differences in policy preferences between these groups can therefore be interpreted as reflecting differences in attribution. Comparing the two *Control* groups, we find no statistically significant differences in the main policy outcomes, although the *Control* & *90* group tends to show lower support for discrimination-related policies (see Table 1.B.12). This difference reaches statistical significance for quotas, with the *Control* & *90* group exhibiting a 0.175 s.d. lower level of support compared to *Control* & *60* ($p < 0.10$; see Table 1.B.13). Taken together, these findings suggest that differences in attribution to women's versus men's wages may partly account for the differences in policy preferences observed between the *Info* & *60* and *Info* & *90* groups.

Turning to the differences between the *Control* and *Info* groups, we find that both the *Info* & *60* and *Info* & *90* treatments significantly increase general support for reducing the gender wage gap (Column (1)). In addition, *Info* & *60* significantly shifts relative support away from child-related toward discrimination-related policies (Column (5)), while simultaneously increasing the preference for prioritizing pay differences related to having children (Column (6)). At this stage, however, it remains unclear how the information provided relates to participants' prior beliefs. To better understand these patterns, the next subsection examines how the treatments interact with individuals' prior beliefs.

1.4.3 Treatment effects by prior

To better understand differences in policy views between the *Control* and *Info* groups, we consider how the provided information relates to individuals' prior beliefs. For individuals with priors less than or equal to 60, assignment to *Info* & *60* provides information weakly above their prior, while assignment to *Info* & *90* provides information strictly above it. In contrast, for individuals with priors above 60, assignment to *Info* & *60* provides information strictly below their prior, whereas assignment to *Info* & *90* provides information weakly above it.[8] We therefore split the sample by whether prior beliefs are less than or equal to 60 or above 60 and rerun equation 1.1 separately for the two subsamples.

Table 1.3. Policy support by treatment status split by prior

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Panel A: Prior $\leq$ 60						
Prior	-0.001	-0.000	-0.001	0.001	-0.003	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Info	0.095	-0.006	-0.116	-0.015	-0.101	0.168*
	(0.075)	(0.074)	(0.074)	(0.063)	(0.067)	(0.086)
Info & 90	0.054	-0.013	0.024	-0.058	0.082	0.290***
	(0.070)	(0.073)	(0.084)	(0.061)	(0.072)	(0.097)
Observations	835	835	835	835	835	835
Control mean	3.99	4.21	-	-	-	2.65
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Prior > 60						
Prior	-0.012**	-0.008	-0.008	-0.005	-0.003	0.005
	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)
Info	0.242**	0.154	-0.092	0.147*	-0.238***	0.045
	(0.105)	(0.111)	(0.102)	(0.086)	(0.085)	(0.129)
Info & 90	-0.179*	-0.167	0.114	-0.204**	0.318***	0.123
	(0.108)	(0.116)	(0.122)	(0.094)	(0.094)	(0.137)
Observations	439	439	439	439	439	439
Control mean	3.90	4.01	-	-	-	2.77
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Panel A reports results for subjects who estimate the percentage of the gender wage gap explained by children to be less than or equal to 60, whereas Panel B reports results for those estimating it to be above 60. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.14. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

8. Only one participant in *Info* & *90* reports a prior greater than 90.

Panel A of Table 1.3 reports results for individuals with priors less than or equal to 60, while Panel B reports results for those with priors above 60. We begin with individuals whose priors are less than or equal to 60. Relative to the *Control* group, the main difference for those in the *Info* treatments concerns preferences to prioritize child-related pay differences. Individuals who receive a weak upward shift exhibit a 0.168 s.d. higher preference relative to the *Control* group ($p = 0.050$). The increase is even larger for individuals receiving a strong upward shift relative to those receiving a weak upward shift (0.290 s.d.; $p < 0.01$). In addition, individuals who are strongly upward shifted display significantly higher general support than those in the *Control* group (0.149 s.d.; $p < 0.05$).

Turning to individuals with priors strictly above 60, those receiving a downward shift exhibit significantly higher general support (0.242 s.d.; $p < 0.05$) and shift their relative support away from child-related policies toward discrimination-related ones (−0.238 s.d.; $p < 0.01$). This change in relative support is more strongly driven by an increase in support for discrimination-related policies (0.147 s.d.; $p < 0.10$). In contrast, individuals receiving an upward shift respond significantly differently from those receiving a downward shift. They exhibit a significantly smaller increase in general support and in support for discrimination-related policies, and a more positive change in their relative support for child-related policies. However, when compared directly to the *Control* group, individuals receiving an upward shift do not exhibit statistically significant differences. The point estimates are nevertheless directionally similar to those observed for individuals with priors less than or equal to 60 who receive a weak upward shift.

These findings help reconcile the earlier result that *Info* & *60* significantly shifts relative support toward discrimination-related policies while simultaneously increasing the preference for prioritizing pay differences related to having children. Participants in *Info* & *60* who were downward shifted reallocate support away from child-related policies toward discrimination-related ones, whereas those who were upward shifted primarily increase their preference for prioritizing pay differences related to children.

As a robustness check, we estimate instrumental variable specifications that exploit variation in actual belief updating rather than treatment assignment alone. Specifically, we instrument the change in beliefs, measured as $Posterior_i - Prior_i$, with treatment assignment. This variable captures both the direction and magnitude of updating, taking positive values for upward revisions and negative values for downward revisions. The IV estimates yield qualitatively similar results to the intention-to-treat effects reported in the main analysis (see Table 1.B.15 for the IV estimates and Table 1.B.16 for the corresponding first stage).

To summarize these findings, we adopt a slightly different approach. Rather than splitting participants by both the information treatment they received and their prior beliefs, we instead group them according to how the provided information relates to their prior. Specifically, we estimate the following specification:

$$Y_i \ = \ \delta_0 + \delta_1 \cdot \mathit{Prior}_i + \delta_2 \cdot \mathit{Info}_i + \delta_3 \cdot \mathit{Info{=}Prior}_i + \delta_4 \cdot \mathit{Info{>}Prior}_i + \delta_5 \cdot \mathbf{X}_i + \mu_i \quad (1.2)$$

where Y denotes the outcome variable capturing the respondent's policy view, *Prior* represents the respondent's initial belief about the share of the gender wage gap explained by children, *Info* is a dummy variable indicating whether the respondent was provided with an empirical estimate of this share, *Info=Prior* is a dummy variable equal to one if the respondent was provided with an empirical estimate equal to their prior, *Info>Prior* equals one if the information exceeds the prior, and **X** includes demographic controls.[9] The coefficient δ_2 measures the difference between the *Control* group and individuals in the *Info* groups who were downward shifted. The coefficient δ_4 measures the difference between individuals in the *Info* groups who were upward shifted and those who were downward shifted, thereby identifying whether treatment effects significantly differ between participants receiving information below versus above their prior. The sum $\delta_2 + \delta_4$ therefore measures the difference between the *Control* group and individuals in the *Info* groups who were upward shifted.[10]

Table 1.4 reports results for equation 1.2. The results show that responses depend significantly on the sign of the information relative to the prior, most notably for relative policy support and support for discrimination-related policies, and to a lesser extent for general support.

Those being downward shifted significantly increase their general support (0.258 s.d.; $p < 0.01$) and significantly reallocate support away from child-related policies and toward discrimination-related policies (−0.249 s.d.; $p < 0.01$). This shift is mainly driven by lower support for paternity leave and higher support for discrimination-related policies, particularly AAP and stricter equal pay legislation (see Table 1.B.17).

Comparing individuals who are upward shifted to those who are downward shifted shows that the former exhibit significantly lower general support (−0.151 s.d.; $p < 0.10$) and significantly lower support for discrimination-related policies (−0.160 s.d.; $p < 0.05$), as well as significantly higher relative support. In particular, they reallocate relatively more support toward child-related policies than toward discrimination-related ones (0.233 s.d.; $p < 0.01$) and exhibit a significantly stronger preference for prioritizing child-related pay differences (0.186 s.d.; $p < 0.10$). Compared to the *Control* group, those being upward shifted increase their general support by only 0.107 s.d. ($p < 0.10$) and significantly increase their preference for prioritizing pay differences related to children by 0.280 s.d. ($p < 0.01$).

9. 661 participants received an empirical estimate above their prior, 153 received an estimate below their prior, and only 37 received an estimate equal to their prior. Given the small sample size of this latter group, we refrain from reporting and interpreting the corresponding results.

10. When reporting differences between the *Control* group and individuals who were upward shifted, we perform a Wald test of the null hypothesis that $\delta_2 + \delta_4 = 0$.

Table 1.4. Policy support by direction of information relative to prior

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Prior	-0.003*	-0.003*	-0.000	-0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Info	0.258***	0.122	-0.124	0.125*	-0.249***	0.095
	(0.088)	(0.092)	(0.089)	(0.073)	(0.077)	(0.107)
Info > Prior	-0.151*	-0.128	0.074	-0.160**	0.233***	0.186*
	(0.087)	(0.092)	(0.092)	(0.073)	(0.078)	(0.110)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.17. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Result 2 (Direction of belief updating) Treatment effects on general support, support for discrimination-related policies, and relative support depend significantly on the direction of the information relative to the prior.

- Those being **downward** shifted significantly increase their general support and significantly reallocate support away from child-related policies and toward discrimination-related ones.
- Those being **upward** shifted also increase their general support, but to a lesser extent, and significantly increase the preference for prioritizing child-related pay differences.

These results suggest that the direction of the shift alone cannot account for the observed treatment effects on general support, as both upward- and downward-shifted participants increase their support. Nor is the direction meaningfully correlated with support for government intervention or child-related policies. To gain a clearer understanding, we next examine heterogeneous effects based on participants' perceptions of the child-related component relative to the non-child-related component of the gender wage gap.

1.4.4 The role of fairness perceptions

Previous research shows that many individuals hold fairness views according to which the acceptance of inequality depends on beliefs about its sources and whether these sources are perceived as fair or unfair (see, e.g., Almås, Hufe, and Weishaar, 2025, for a review). In the context of the gender wage gap, individuals may hold different beliefs about the sources and fairness of child-related pay differences as well as of pay differences unrelated to children. Shifting beliefs about the contribution of children to the gender wage gap alters the perceived importance of child-related factors relative to other determinants. As a result, the effect of such shifts on support for reducing the gender wage gap might depend on how fair an individual perceives child-related pay differences to be compared to other explanatory factors.

To capture these perceptions, we elicit participants' views on the fairness of the child-related part relative to the remaining part of the gap on a 5-point Likert scale ranging from 1 (*"much less fair"*) to 5 (*"much more fair"*). Since both *Info* groups received information about the size of the child-related component prior to answering this question, we can test whether information affects perceived relative fairness. Table 1.B.18 shows mean perceptions about relative fairness by treatment status. Tests of equality of means show no statistically significant differences, suggesting that these perceptions are independent of beliefs about children's contribution to the gender wage gap.[11] Across the full sample, the child-related part is perceived as slightly less fair on average, with considerable heterogeneity across participants (see Figure 1.B.2).

We now examine whether the impact of information on policy support depends on individuals' fairness perceptions. Specifically, we test how participants who view the child-related component as less fair and those who view it as equally or more fair respond to information above or below their prior, and whether these responses differ between the two groups. To do this, we introduce a binary variable, *Child less fair*, which equals 1 if a participant perceives the child-related part as strictly less fair than the part not related to children. We then rerun equation 1.2 separately for the two subsamples. In addition, we estimate a pooled specification interacting all covariates with *Child less fair*, which allows for a direct test of whether the coefficients differ across the two groups.

Panel A of Table 1.5 reports results for participants who perceive the child-related part as equally or more fair, while Panel B reports results for those who perceive it as less fair. Table 1.B.19 reports the corresponding pooled interaction

11. Perceptions do not differ between the pooled *Control* group and the two *Info* groups ($p = 0.610$ and $p = 0.187$, respectively), nor between *Info & 60* and *Info & 90* ($p = 0.408$).

Table 1.5. Policy support by direction of information relative to prior split by fairness perceptions

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Panel A: Equally or more fair						
Prior	-0.006**	-0.005**	-0.002	-0.004*	0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Info	0.367***	0.306**	-0.114	0.242**	-0.356***	0.111
	(0.135)	(0.131)	(0.140)	(0.107)	(0.121)	(0.136)
Info > Prior	-0.359***	-0.371***	-0.066	-0.305***	0.240*	0.017
	(0.138)	(0.135)	(0.149)	(0.110)	(0.123)	(0.141)
Observations	641	641	641	641	641	641
Control mean	3.79	3.94	-	-	-	2.64
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Less fair						
Prior	0.002	0.001	0.003***	0.002	0.002	0.005*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Info	0.152	-0.082	-0.143*	-0.009	-0.134	0.112
	(0.103)	(0.127)	(0.085)	(0.088)	(0.088)	(0.171)
Info > Prior	0.018	0.103	0.174**	-0.024	0.198**	0.299*
	(0.097)	(0.121)	(0.085)	(0.085)	(0.090)	(0.168)
Observations	633	633	633	633	633	633
Control mean	4.16	4.38	-	-	-	2.75
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. Panel A reports results for respondents who believe that the child-related part is equally or more fair, whereas Panel B reports results for those who believe it is less fair than the part not explained by children. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.20. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

specification. Two key findings emerge. First, individuals who perceive the child-related component as equally or more fair and those who view it as less fair respond similarly in terms of relative support for child-related versus discrimination-related policies, although the underlying shifts in support for specific policies differ across groups.

Among participants who perceive the child-related component as equally or more fair, the direction of belief updating is negatively related to support for discrimination-related policies and therefore positively related to relative support for child-related versus discrimination-related policies. Downward-shifted individuals increase their support for discrimination-related policies relative to the *Control* group (0.242 s.d.; $p < 0.05$), driven by higher support for both quotas and AAP (see Table 1.B.20). This leads to a significant reallocation of support away from child-related toward discrimination-related policies (−0.356 s.d.; $p < 0.01$). By contrast, upward-shifted individuals show only one significant difference relative to the *Control* group, namely reduced support for child-related policies (−0.179 s.d.; $p < 0.10$), mainly driven by lower support for child care subsidies (see Table 1.B.20).

Among individuals who perceive the child-related component as less fair, the relationship between the direction of the signal and support for both child-related and discrimination-related policies is qualitatively more positive than among those who perceive it as equally or more fair. This pattern is visible across all specific policies, with the largest differences by fairness perception observed for support for quotas, followed by child care subsidies (see Table 1.B.21). As a result, among participants who perceive the child-related component as less fair, the direction of the information relative to the prior is positively related to relative support, mainly through a positive relationship with support for child-related policies. Downward-shifted individuals exhibit a decrease in support for child-related policies (−0.143 s.d.; $p < 0.10$), driven predominantly by lower support for paternity leave relative to the *Control* group (see Table 1.B.20). By contrast, upward-shifted individuals respond more positively in their support for child-related policies, although they do not significantly change their support relative to the *Control* group. In addition, upward-shifted individuals exhibit a stronger preference for prioritizing child-related pay differences relative to the *Control* group (0.411 s.d.; $p < 0.01$).

The second main finding is that the effect of shifts in beliefs about the contribution of children to the gender wage gap on overall support depends significantly on whether individuals perceive the child-related component as more or less fair than the part not explained by children. Participants who perceive the child-related component as less fair exhibit a more positive relationship between overall support and the direction of the information signal.

Among participants who perceive the child-related component as equally or more fair, we find a strong negative relationship between the direction of the information relative to the prior and overall support. Downward-shifted individuals significantly increase their overall support, reflected in both general support (0.367 s.d.; $p < 0.01$) and support for government intervention (0.306 s.d.; $p < 0.05$). By contrast, upward-shifted individuals do not significantly change their overall support relative to the *Control* group.

Turning to participants who perceive the child-related component as less fair, the direction of the information relative to the prior is positively, but not significantly, related to overall support. Relative to the *Control* group, only upward-shifted individuals exhibit a significant increase in general support (0.170 s.d.; $p < 0.05$).

To summarize, participants increase general support when information indicates that the component of the gender wage gap they perceive as less fair is larger, whereas they do not reduce support when information suggests that the component they perceive as more fair is larger. This asymmetry helps reconcile the earlier finding that both upward- and downward-shifted participants increase general support on average.

Result 3 (Fairness) The relationship between the direction of belief updating and overall support as well as support for specific policies depends significantly on whether individuals view the child-related component as more or less fair than the part not explained by children.

a. For individuals who perceive the child-related component as **equally or more fair**, the direction of belief updating is negatively related to overall support and support for discrimination-related policies, resulting in a positive relationship with relative support for child-related policies.
b. For individuals who perceive the child-related component as **less fair**, the direction of belief updating is positively related to support for child-related policies, relative support for child-related policies, and the preference for prioritizing child-related pay differences.

A similar analysis based on whether participants perceive the child-related part as more or less efficient is presented in Appendix 1.B.2.4. In contrast to fairness perceptions, we do not observe statistically significant differences in responses by efficiency perceptions. This is in line with previous findings showing that efficiency concerns are less predictive of the acceptance of inequality and support for redistributive policies than fairness concerns (e.g., Stantcheva, 2021; Almås et al., 2024).

1.4.4.1 Beliefs about sources of gender wage gap

In this subsection, we examine how beliefs about the role of children in the gender wage gap and perceptions of the relative fairness of the child-related component jointly shape policy views. Specifically, we test whether differences in policy responses across fairness perceptions operate through beliefs about the underlying sources of the gender wage gap. These sources include individual choice, as well as more specific channels such as discrimination and differences in personal traits.

As a first step, Table 1.B.25 examines how the direction of the shift in beliefs about the contribution of children to the gender wage gap affects beliefs about these sources, depending on individuals' fairness perceptions. The results show that the effect of the shift on beliefs about the sources of the gap differs significantly by fairness perceptions. Individuals who perceive the child-related component as less fair exhibit a significantly more negative relationship between the direction of the shift and the belief that the gender wage gap reflects individual choice, and a significantly more positive relationship with the belief that the gap reflects discrimination. This pattern suggests that individuals with different fairness perceptions hold systematically different views about whether the child-related and non-child-related components of the gap are reflecting individual choice or discrimination.

In a second step, Table 1.B.26 examines whether differences in the impact of information on policy support across fairness perceptions persist once beliefs about these sources are controlled for. The results show that controlling for beliefs about the sources of the gender wage gap accounts for the heterogeneity in policy responses, suggesting that the joint effect of beliefs about the role of children and fairness perceptions operates through beliefs about these sources.

1.5 Conclusion

This paper studies beliefs about the role of children in driving the gender wage gap and how these beliefs shape policy preferences. We document that beliefs about the share of the gender wage gap explained by having children are highly dispersed and, on average, substantially lower than the empirical estimates. Leveraging randomized information treatments that exogenously shift beliefs about the importance of children in explaining the gap, we show that increasing the perceived contribution of children increases the preference for prioritizing pay differences related to children and leads to a significant reallocation of support across policy types. In particular, higher perceived contributions of children shift support toward child-related policies, primarily through higher support for maternity leave, and away from discrimination-related policies such as quotas and affirmative action programs.

We further show that treatment effects, particularly on support for discrimination-related policies and relative support, depend on the direction of belief updating. Individuals whose beliefs are shifted downward reallocate support away from child-related policies toward discrimination-related policies, while individuals whose beliefs are shifted upward increase their preference for prioritizing pay differences related to children. Finally, we demonstrate that fairness perceptions are central to these responses, particularly for overall support. Participants who perceive the child-related component as less fair exhibit a more positive relationship between the direction of belief updating and overall support. More broadly, information increases overall policy support only when it raises the perceived importance of the component of the wage gap individuals view as less fair.

While prior work has highlighted the empirical importance of the child penalty, our study shows that this central driver of gender inequality is largely underestimated and that correcting this informational gap has meaningful effects. We provide novel causal evidence that beliefs about specific explanations for inequality matter for how individuals allocate support across competing policies.

Our methodology and findings open possibilities for future research. We fix beliefs about the size of an inequality and provide information about a percentage explained by a specific factor. This methodology can be applied to other types of inequality. Furthermore, while we focus on beliefs about the role of children in driving the gender wage gap, other contributing factors, such as discrimination, occupational sorting, or differences in negotiation behavior, may be subject to similar misperceptions. Exploring beliefs about these factors, and how they interact with fairness and efficiency considerations, could yield further insights into the relationship between narratives and policy preferences. By introducing the concepts of relative support and relative fairness, we initiate a discussion about trade-offs in such preferences and perceptions. Moreover, we address the question of how relative fairness operates and set the stage for further investigation of what determines it.

These directions build upon the central contribution of this paper: that beliefs about the causes of inequality are consequential, and understanding them is critical for designing effective policy interventions.

References

Alesina, Alberto, Matteo F Ferroni, and Stefanie Stantcheva. 2021. "Perceptions of racial gaps, their causes, and ways to reduce them." Working Paper 29245. National Bureau of Economic Research. [11]

Alesina, Alberto, Stefanie Stantcheva, and Edoardo Teso. 2018. "Intergenerational mobility and preferences for redistribution." *American Economic Review* 108 (2): 521–54. [11]

Almås, Ingvild, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden. 2024. "Attitudes to inequality: preferences and beliefs." *Oxford Open Economics* 3 (Supplement_1): i64–i79. [35]

Almås, Ingvild, Paul Hufe, and Daniel Weishaar. 2025. "Experimental evidence on attitudes toward inequality and fairness." *Annual Review of Economics* 17. [12, 32]

Becker, Sascha O., Ana Fernandes, and Doris Weichselbaumer. 2019. "Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment." *Labour Economics* 59 (C): 139–52. [11]

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics* 2 (3): 228–55. [11]

Boinet, Césarine, Jonathan Norris, Agnese Romiti, Zhan Shi, and Paul Telemo. 2024. "Beliefs on children's human capital formation and mothers at work." Working Paper 17574. IZA Discussion Paper Series. [11]

Boneva, Teodora, Marta Golin, Katja Kaufmann, and Christopher Rauh. 2026. "Beliefs about maternal labour supply." *Economic Journal* 136 (674): 373–401. [11]

Caplin, Andrew, SÃ¸ren Leth-Petersen, and Christopher Tonetti. 2025. "Beliefs and realities of work and childcare after childbirth." Working Paper 34289. National Bureau of Economic Research. [11]

Casarico, Alessandra, Jana Schuetz, and Silke Uebelmesser. 2024. "Gender inequality over the life cycle, information provision and policy preferences." Working Paper 10916. CESifo Working Paper Series. [11]

Correll, Shelley J., Stephen Benard, and In Paik. 2007. "Getting a job: Is there a motherhood penalty?" *American Journal of Sociology* 112 (5): 1297–339. [11]

Cortés, Patricia, and Jessica Pan. 2023. "Children and the remaining gender gaps in the labor market." *Journal of Economic Literature* 61 (4): 1359–409. [7, 9, 11, 13, 15, 16, 42]

Costa-Ramón, Ana, Ursina Schaede, Michaela Slotwinski, and Anne Ardila Brenoe. 2024. "(Not) thinking about the future: Inattention and maternal labor supply." Working Paper 11359. CESifo Working Paper Series. [11]

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American Economic Review* 104 (4): 1091–119. [11]

Goldin, Claudia. 2024. "Nobel lecture: An evolving economic force." *American Economic Review* 114 (6): 1515–39. [7, 11]

Gong, Yifan, Ralph Stinebrickner, and Todd Stinebrickner. 2022. "Marriage, children, and labor supply: Beliefs and outcomes." *Journal of Econometrics* 231 (1): 148–64. Annals Issue: Subjective Expectations & Probabilities in Economics. [11]

Günther, Isabel, and Bruno Martorano. 2025. "Inequality, social mobility and redistributive preferences." *Journal of Economic Behavior & Organization* 229: 106828. [11]

Haaland, Ingar, and Christopher Roth. 2023. "Beliefs about racial discrimination and support for pro-black policies." *Review of Economics and Statistics* 105 (1): 40–53. [11]

Henkel, Aljosha, Ernst Fehr, Julien Senn, and Thomas Epper. 2025. "Beliefs about inequality and the nature of support for redistribution." *Journal of Public Economics* 246: 105350. [11]

Hotz, V Joseph, Per Johansson, and Arizo Karimi. 2018. "Parenthood, family friendly workplaces, and the gender gaps in early work careers." Working paper 24173. National Bureau of Economic Research. [11]

Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante. 2025. "The child penalty atlas." *Review of Economic Studies* 92 (5): 3174–207. [11]

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11 (4): 181–209. [7, 11]

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. 2015. "How elastic are references for redistribution? Evidence from randomized survey experiments." *American Economic Review* 105 (4): 1478–508. [11]

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington. 2018. "The mommy effect: Do women anticipate the employment effects of motherhood?" Working Paper 24740. National Bureau of Economic Research. [11, 22]

Olivetti, Claudia, Jessica Pan, and Barbara Petrongolo. 2024. "The evolution of gender in the labor market." In *Handbook of Labor Economics,* 5: 619–77. Elsevier. [7, 11]

Settele, Sonja. 2022. "How do beliefs about the gender wage gap affect the demand for public policy?" *American Economic Journal: Economic Policy* 14 (2): 475–508. [11]

Stantcheva, Stefanie. 2021. "Understanding tax policy: How do people reason?" *Quarterly Journal of Economics* 136 (4): 2309–69. [35]

Appendix 1.A Experiment

1.A.1 Treatment conditions: Design

Panel A: *Low estimate: 60*

Introduction: Gender wage gap and children

When people **have their first child**, they may **change their careers** or **face limitations** that affect their work. These changes can impact their wages.

Having a child may impact women's and men's wages differently and the impact can be measured separately for each. To do this, **mothers** are compared to **women without children**, and **fathers** to **men without children**, while accounting for factors like education, age, race and region.

Research shows that men's wages do not fall after having a child. You should assume that **children have no effect on men's wages**. However, **women's wages tend to fall**. This **increases the gap** between women's and men's wages and can therefore help **explain the gender wage gap**.

- The **larger the drop** in women's wages compared to men's, **the more children explain** the gender wage gap.
- The **smaller the drop**, the **less children explain** the gender wage gap.

On the next page, you'll guess **what percentage of the gender wage gap** is **explained by children**. The remaining percentage exists even in the absence of children and is therefore caused by other factors.

Panel B: *High estimate: 90*

Introduction: Gender wage gap and children

When people **have their first child**, they may **change their careers** or **face limitations** that affect their work. These changes can impact their wages.

Having a child may impact women's and men's wages differently and the impact can be measured separately for each. To do this, **mothers** are compared to **women without children**, and **fathers** to **men without children**, while accounting for factors like education, age, race and region.

Research shows that men's wages do not fall after having a child. You should assume that **children have a slightly positive effect on men's wages**. However, **women's wages tend to fall**. This **increases the gap** between women's and men's wages and can therefore help **explain the gender wage gap**.

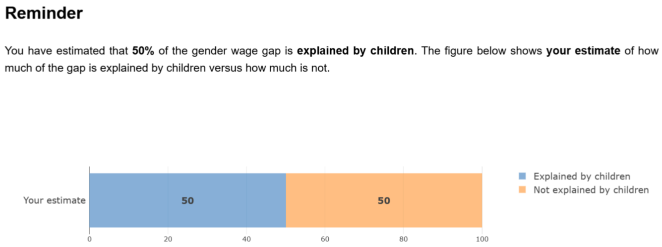
- The **larger the drop** in women's wages compared to men's, **the more children explain** the gender wage gap.
- The **smaller the drop**, the **less children explain** the gender wage gap.

On the next page, you'll guess **what percentage of the gender wage gap** is **explained by children**. The remaining percentage exists even in the absence of children and is therefore caused by other factors.

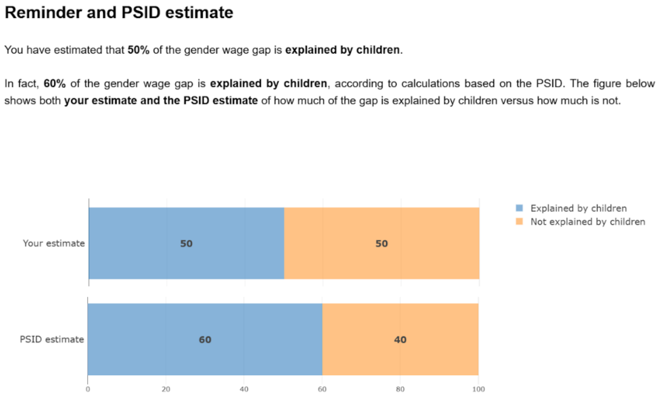
Notes: This figure shows the introduction screen presented to participants, explaining the percentage of the gender wage gap attributed to children. Panel A corresponds to the *Low estimate: 60* treatment and Panel B to the *High estimate: 90* treatment.

Figure 1.A.1. Treatment groups: *Low estimate: 60* vs. *High estimate: 90*

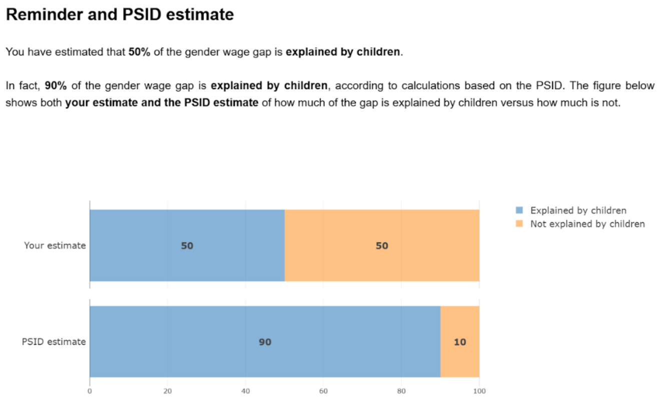
Panel A: *Control* & *60* and *Control* & *90*



Panel B: *Info* & *60*



Panel C: *Info* & *90*



Notes: This figure shows the main variation across treatment groups. Participants in the *Info* groups are presented with the empirically estimated percentage based on the PSID and are reminded of their own prior estimate, whereas participants in the *Control* groups are only reminded of their own prior estimate. Panel A corresponds to the *Control* treatments, Panel B to the *Info* & *60* treatment and Panel C to the *Info* & *90* treatment.

Figure 1.A.2. Treatment groups: *Control* vs. *Info*

1.A.2 Treatment conditions: Calculation of the PSID estimates

Cortés and Pan (2023) perform a decomposition of the gender wage gap into a child-related component and a component unrelated to children. They report that 90 percent of the gender wage gap is explained by having children.

To obtain a second estimate, we slightly modified the definition of the child-related part of the gender wage gap to include only the effect of children on women's wages. To compute the corresponding percentage, we repeated the same procedure, with the only change being the exclusion of the effect of children on men's wages from the child-related part. If men's wages were truly unaffected by children, this would yield the same estimate. However, because men's wages slightly increase after having children, the resulting share of the gender wage gap explained by children under this modified definition is 60 percent.

To make the two treatments as comparable as possible, we defined the percentage explained by children in exactly the same way, but instructed participants in *Info & 60* to assume that men's wages are unaffected by children, so they would consider only the effect of children on women's wages when estimating the percentage. Participants in *Info & 90* were instead instructed to assume that men's wages slightly increase after having children.

Below, we present the steps of the decomposition in more detail.

Step 1: For men and women separately, the following regression is estimated:

$$Y^g_{ip\tau} = \sum_y \sum_{\tau=-5}^{10} \beta^g_{y\tau} \cdot \mathbb{1}[\tau = t - e^i] \cdot \mathbb{1}[y = p] + \sum_k \alpha^g_k X^g_{kit} + \nu^g_{ipt} \tag{1.A.1}$$

where $Y^g_{ip\tau}$ denotes the log wage of individual i at event-time τ and time period p. The resulting coefficients $\hat{\beta}^g_{p\tau}$ estimate the effect of children on wages for women and men at each event-time τ and time period p.

Step 2: The gender wage gap, defined as $\Delta_p = \mathbb{E}\left[Y^m_{ip\tau}\middle|p,m\right] - \mathbb{E}\left[Y^w_{ip\tau}\middle|p,f\right]$, can be decomposed as follows:

$$\Delta_p = \mathbb{E}\left[\hat{Y}^m_{ip\tau}\middle|p,m\right] - \mathbb{E}\left[\hat{Y}^f_{ip\tau} - (\hat{\beta}^f_{p\tau} - \hat{\beta}^m_{p\tau})\middle|p,f\right] + \mathbb{E}\left[\hat{\beta}^m_{p\tau} - \hat{\beta}^f_{p\tau}\middle|p,f\right] \tag{1.A.2}$$

The first term represents the expected predicted wages for men. The second term reflects the expected predicted wages for women, assuming the impact of children on their wages is the same as for men. The third term captures the child penalty—i.e., the average differential impact of children on women's wages relative to men's. This third term is defined as the child-related part of the gender wage gap.

The percentage of the wage gap explained by children is then calculated by dividing this third term by the total gender wage gap.

Modification: For the modification, we used the following decomposition and defined the third term as the child-related part.

$$\Delta_p = \mathbb{E}\left[\hat{Y}^m_{ip\tau}\middle|p,m\right] - \mathbb{E}\left[\hat{Y}^f_{ip\tau} - \hat{\beta}^f_{p\tau}\middle|p,f\right] + \mathbb{E}\left[-\hat{\beta}^f_{p\tau}\middle|p,f\right] \tag{1.A.3}$$

In the data, children have a slightly positive effect on men's wages. If men's wages were unaffected by children, such that $\hat{\beta}^m_{p\tau} = 0$, then the third term in this decomposition would coincide with the child-related part in the original specification.

1.A.3 Variables

Table 1.A.1. Main variables

Category	**Variable**	***Question***	**Format**
Prior beliefs	Size: Gender wage gap	*How many dollars do you think a woman earns for every $100 earned by a man?*	Slider ($0–$200)
	Prior	*What percentage of the gender wage gap do you think is explained by children?*	Slider (0–100%)
Posterior beliefs	Posterior	*Let us ask you again, what percentage of the gender wage gap do you think is explained by children?*	Slider (0–100%)
	Subjective child importance	*On a scale from 0 to 10, how important are children in explaining the wage gap?*	11-point scale
Overall support	General support	*How important do you believe it is for society to work toward reducing the gender wage gap?*	5-point Likert scale
	Government action support	*To what extent do you support government action to reduce the wage gap?*	5-point Likert scale
Specific support	Policies	*Please indicate your level of support:* Maternity leave, paternity leave, child care subsidy, equal pay legislation, affirmative action, gender quotas, abortion rights, infrastructure, defense, healthcare	5-point Likert scale
	Child policies	Generated as the average support for the following policies: Maternity leave, paternity leave, child care subsidy	
	Discrimination policies	Generated as the average support for the following policies: Equal pay legislation, affirmative action, gender quotas	
	Unrelated policies	Generated as the average support for the following policies: Infrastructure, defense, healthcare	
Relative support	Prioritization child-related	*Do you think the government should prioritize reducing gender pay differences caused by children or those caused by other factors?*	5-point Likert scale
	Child-Discrimination	Generated as the difference between the support of child and discrimination policies	
Donation	Donation	*Please enter the amount you would like to donate to CCAoA.*	Text box (0–100)
Perceptions about the child-related part	Relative fairness	*Do you think it is more or less fair when pay differences between women and men are caused by having children, compared to when they are caused by other factors?*	5-point Likert scale
	Relative efficiency	*Do you think that pay differences between women and men caused by having children reflect a more or less efficient use of talent, meaning more or less result from women and men just being better suited to different kinds of tasks and work, than pay differences caused by other factors?*	5-point Likert scale

Notes: This table reports the main variables used in the analysis. For each variable, the table presents the variable name, the exact survey question, and the response format.

Table 1.A.2. Additional variables

Category	Variable	*Question*	Format
Beliefs and perceptions about	Fairness	*How (un)fair do you find the gender wage gap?*	5-point Likert scale
the overall gender wage gap	Efficiency	*To what extent do you agree that the gender wage gap reflects an efficient use of talent, meaning it results from women and men just being better suited to different kinds of tasks and work?*	5-point Likert scale
	Sources	*To what extent do you agree that the gender wage gap is a result of the following sources of inequality?* (choice, discrimination, preferences, skills, or effort)	5-point Likert scale
	Biological vs. societal attribution	*For each of the following differences, how much do you think it is influenced by biological rather than societal factors?* (hourly wages, preferences, skills, performance, and effort)	6-point Likert scale (incl. answer: *There is no such difference*)
Understanding of information	Understanding: child-related part	*Given the figure on the previous page, what percentage of the gender wage gap is explained by children?*	Slider (0–100%)
	Understanding: non-child-related part	*Given the figure on the previous page, what percentage of the gender wage gap is not explained by children?*	Slider (0–100%)
Additional variables	Demographics	Gender, political party, age, education, own income, marital status, having children, religion, ethnicity, employment status, (voting behavior 2024, state, household income, occupation)	Drop-down multiple choice menus
	Trust in government	*How often do you think you can trust the government to do what is right?*	5-point Likert scale
	Norms	*To what extent do you agree that reducing the gender wage gap will inevitably reduce earnings or opportunities for some individuals or groups?*	5-point Likert scale
	Zero-sum	*To what extent do you agree that men and women should adhere to traditional gender norms, e.g., men as breadwinners, women as homemakers?*	5-point Likert scale

Notes: This table displays the additional variables used in the analysis. For each variable, the table presents the variable name, the exact survey question, and the response format.

1.A.4 Screenshots of the experiment

The experiment was conducted in English using the U.S. sample of the platform Prolific. It was programmed in Qualtrics. The screenshots of the experiment can be found in the digital appendix on the bonndoc publication server:
https://nbn-resolving.org/urn:nbn:de:hbz:5-91417

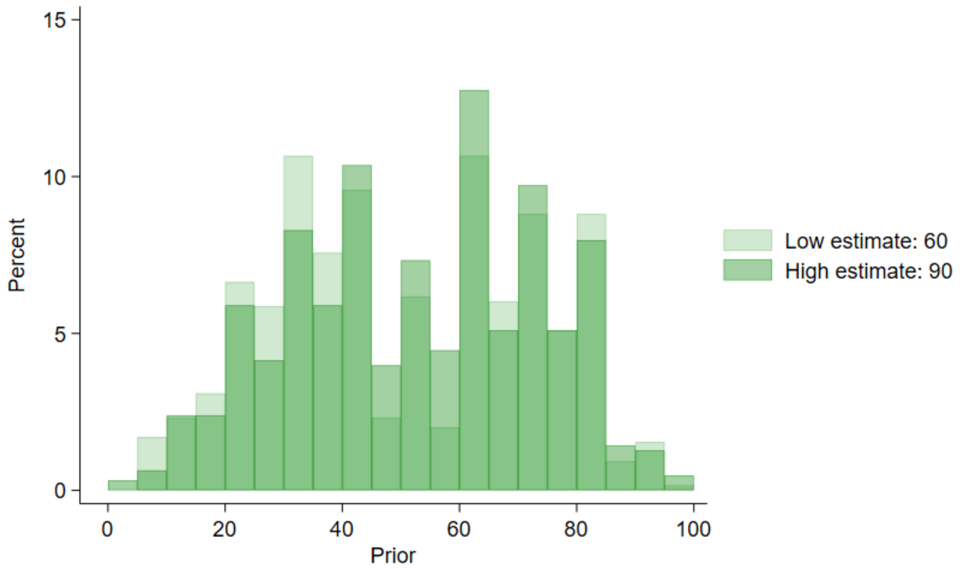
Appendix 1.B Additional tables and figures

1.B.1 Descriptive evidence

Table 1.B.1. Demographics by treatment groups used for the analysis

	Treatment group:		
	Control	Info & 60	Info & 90
N	423 (33.2%)	441 (34.6%)	410 (32.2%)
Gender			
Male	195 (46.1%)	224 (50.8%)	212 (51.7%)
Female	226 (53.4%)	212 (48.1%)	197 (48.0%)
Other gender	2 (0.5%)	5 (1.1%)	1 (0.2%)
Political party			
Republican	172 (40.7%)	178 (40.4%)	150 (36.6%)
Democrat	148 (35.0%)	168 (38.1%)	160 (39.0%)
Independent	88 (20.8%)	77 (17.5%)	90 (22.0%)
Other Party	15 (3.5%)	18 (4.1%)	10 (2.4%)
Age	40.125 (13.164)	39.522 (12.697)	40.105 (13.138)
College degree			
No college degree	98 (23.2%)	105 (23.8%)	85 (20.7%)
College degree	325 (76.8%)	336 (76.2%)	325 (79.3%)
Own income			
Less than 25k	101 (23.9%)	92 (20.9%)	83 (20.2%)
Between 25k and 50k	82 (19.4%)	96 (21.8%)	94 (22.9%)
Between 50k and 75k	105 (24.8%)	98 (22.2%)	102 (24.9%)
Between 75k and 100k	70 (16.5%)	77 (17.5%)	59 (14.4%)
Above 100k	65 (15.4%)	78 (17.7%)	72 (17.6%)
Marital status			
Not married	200 (47.3%)	224 (50.8%)	183 (44.6%)
Currently married	223 (52.7%)	217 (49.2%)	227 (55.4%)
Has children			
No children	164 (38.8%)	180 (40.8%)	162 (39.5%)
Has children	259 (61.2%)	261 (59.2%)	248 (60.5%)
Religion			
Christianity	277 (65.5%)	292 (66.2%)	263 (64.1%)
Other Religion	17 (4.0%)	18 (4.1%)	13 (3.2%)
Atheism	83 (19.6%)	88 (20.0%)	93 (22.7%)
None of the above	46 (10.9%)	43 (9.8%)	41 (10.0%)
Ethnicity			
White	287 (67.8%)	298 (67.6%)	265 (64.6%)
Black	95 (22.5%)	90 (20.4%)	86 (21.0%)
Other Ethnicity	41 (9.7%)	53 (12.0%)	59 (14.4%)
Employment status			
Not employed	74 (17.5%)	56 (12.7%)	56 (13.7%)
Employed	349 (82.5%)	385 (87.3%)	354 (86.3%)

Notes: This table reports the main demographic characteristics used as controls in the subsequent analysis by treatment groups used for the analysis. The number of observations for each category is reported, with the corresponding share of the sample displayed in parentheses.



Notes: This figure shows the distribution of responses to the question of what percentage of the gender wage gap respondents believe is explained by children separately for *Low estimate: 60* and *High estimate: 90*. Responses were provided using a slider ranging from 0% to 100%.

Figure 1.B.1. Distribution of prior beliefs split by *Low estimate: 60* and *High estimate: 90*

Table 1.B.2. Priors by demographics

	(1) Size of gender wage gap	(2) Percent explained by children
Female	-3.261***	2.173*
	(0.856)	(1.234)
Democrat	-0.286	-2.704*
	(1.086)	(1.487)
Independent	0.837	-1.018
	(1.174)	(1.832)
Age	-0.185	-0.130
	(0.189)	(0.284)
Age^2	0.002	0.003
	(0.002)	(0.003)
College degree	0.521	-0.502
	(1.055)	(1.615)
Income	0.035	-0.041
	(0.276)	(0.437)
Currently married	0.813	-0.159
	(1.026)	(1.540)
Has children	-0.816	4.430***
	(0.963)	(1.614)
Atheism	1.120	1.006
	(0.867)	(1.746)
Black	-1.090	1.836
	(1.166)	(1.523)
Other Ethnicity	-0.640	3.867**
	(1.348)	(1.949)
Employed	-0.264	0.563
	(1.151)	(2.054)
Observations	1274	1274

Notes: This table reports coefficient estimates from a linear regression model. See Table 1.B.1 for the corresponding baseline categories. Coefficients are reported only for categories representing more than 10% of the sample. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.3. Priors by *Control* groups

	Treatment group	
	Control & 60	Control & 90
Prior	48.325	50.424
	(22.095)	(21.905)
Size gender wage gap	79.029	79.834
	(12.810)	(14.806)
Observations	206	217

Notes: This table reports mean prior beliefs about both the size of the gender wage gap and the share of the gender wage gap explained by children for the two treatment groups *Control* & *60* and *Control* & *90*.

1.B.2 Results

1.B.2.1 Policy support by treatment status

Table 1.B.4. Support for specific policies by treatment status

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Prior	-0.001	-0.001	-0.001	-0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Info	-0.128*	-0.109	-0.071	-0.012	0.073	0.074
	(0.074)	(0.070)	(0.070)	(0.060)	(0.061)	(0.062)
Info & 90	0.148*	0.038	-0.019	-0.102*	-0.118**	-0.093
	(0.081)	(0.077)	(0.076)	(0.061)	(0.060)	(0.063)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.5. Donation child care

	(1) Donation child care	(2) Donation child care	(3) Donation child care
Prior	0.002	0.001	0.001
	(0.001)	(0.001)	(0.002)
Info	0.059	0.140	0.119
	(0.064)	(0.100)	(0.137)
Info & 90	-0.094		
	(0.063)		
Info > Prior		-0.163*	-0.217
		(0.097)	(0.134)
Child less fair			-0.030
			(0.161)
Child less fair × Prior			0.000
			(0.003)
Child less fair × Info			0.042
			(0.199)
Child less fair × Info > Prior			0.101
			(0.195)
Observations	1274	1274	1274
Control mean	26.26	26.26	26.26
Demographics	Yes	Yes	Yes

Notes: Column (1) reports coefficient estimates from linear regressions estimating equation 1.1, column (2) reports estimates based on equation 1.2, and column (3) reports estimates based on equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less fair than the part unrelated to children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in the second specification and the dummy variables *Info=Prior* and *Child less fair×Info=Prior* are included in the third specification, but omitted from the table due to the small number of observations. The dependent variable measures the amount, out of $100, donated to Child Care Aware of America (CCAoA). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.6. Policy support by treatment status interacted with gender

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Female	0.588***	0.544***	-0.000	0.526***	-0.526***	0.001
	(0.156)	(0.151)	(0.147)	(0.129)	(0.134)	(0.173)
Prior	-0.001	-0.001	-0.003	0.000	-0.004*	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Prior	-0.002	-0.002	0.004*	-0.002	0.006***	0.002
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	0.241**	0.169*	-0.036	0.135*	-0.172**	0.161*
	(0.098)	(0.097)	(0.093)	(0.082)	(0.080)	(0.098)
Info & 90	-0.032	-0.107	0.004	-0.114	0.118	0.401***
	(0.088)	(0.090)	(0.099)	(0.075)	(0.084)	(0.106)
Female × Info	-0.175	-0.214*	-0.125	-0.167	0.042	-0.041
	(0.122)	(0.123)	(0.120)	(0.102)	(0.106)	(0.142)
Female × Info & 90	0.005	0.068	0.100	0.006	0.094	-0.368**
	(0.115)	(0.121)	(0.132)	(0.100)	(0.113)	(0.158)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with gender. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.7. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.7. Support for specific policies by treatment status interacted with gender

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Female	0.131	-0.208	0.076	0.571***	0.518***	0.487***
	(0.181)	(0.169)	(0.167)	(0.149)	(0.153)	(0.159)
Prior	-0.003	-0.004*	-0.002	-0.000	0.001	-0.000
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Female × Prior	0.005	0.006**	0.003	-0.002	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	-0.054	-0.113	0.058	0.080	0.186*	0.141
	(0.118)	(0.103)	(0.103)	(0.093)	(0.095)	(0.103)
Info & 90	0.112	0.020	-0.118	-0.123	-0.125	-0.093
	(0.122)	(0.109)	(0.109)	(0.086)	(0.087)	(0.095)
Female × Info	-0.141	0.013	-0.247*	-0.171	-0.210*	-0.120
	(0.149)	(0.139)	(0.139)	(0.121)	(0.122)	(0.124)
Female × Info & 90	0.071	0.033	0.196	0.031	0.002	-0.014
	(0.160)	(0.152)	(0.149)	(0.121)	(0.119)	(0.123)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with gender. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.8. Policy support by treatment status interacted with parental status

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Has children	0.184	0.356**	0.217	0.270**	-0.053	0.046
	(0.163)	(0.160)	(0.157)	(0.134)	(0.143)	(0.185)
Prior	-0.001	-0.001	0.001	0.001	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Has children × Prior	-0.001	-0.002	-0.003	-0.002	-0.001	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	0.219**	0.176*	-0.189*	0.101	-0.290***	0.188*
	(0.097)	(0.097)	(0.098)	(0.082)	(0.089)	(0.106)
Info & 90	-0.060	-0.169*	-0.047	-0.174**	0.127	0.257**
	(0.095)	(0.097)	(0.120)	(0.081)	(0.101)	(0.111)
Has children × Info	-0.115	-0.201	0.137	-0.095	0.232**	-0.092
	(0.124)	(0.125)	(0.124)	(0.104)	(0.110)	(0.143)
Has children × Info & 90	0.058	0.170	0.175	0.119	0.056	-0.049
	(0.121)	(0.126)	(0.143)	(0.103)	(0.122)	(0.155)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with whether the participant has children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.9. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.9. Support for specific policies by treatment status interacted with parental status

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Has children	0.170	0.104	0.376**	0.292*	0.280*	0.237
	(0.195)	(0.177)	(0.177)	(0.155)	(0.163)	(0.162)
Prior	0.001	0.000	0.001	-0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Has children × Prior	-0.004	-0.003	-0.003	0.001	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	-0.246**	-0.258**	-0.062	0.085	0.093	0.125
	(0.121)	(0.111)	(0.114)	(0.096)	(0.099)	(0.099)
Info & 90	0.079	-0.020	-0.199	-0.167*	-0.171*	-0.184*
	(0.145)	(0.128)	(0.129)	(0.094)	(0.096)	(0.101)
Has children × Info	0.190	0.242*	-0.021	-0.161	-0.036	-0.088
	(0.153)	(0.143)	(0.143)	(0.123)	(0.126)	(0.127)
Has children × Info & 90	0.120	0.100	0.304*	0.108	0.093	0.155
	(0.173)	(0.158)	(0.158)	(0.123)	(0.122)	(0.127)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with whether the participant has children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.10. Policy support by treatment status interacted with party affiliation

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Democrat	0.206	0.218	0.170	0.164	0.006	-0.398**
	(0.151)	(0.140)	(0.139)	(0.123)	(0.133)	(0.174)
Prior	-0.003*	-0.004**	-0.002	-0.002	0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Democrat × Prior	0.004*	0.005**	0.003	0.004**	-0.001	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Info	0.178**	0.124	-0.114	0.033	-0.147**	0.070
	(0.080)	(0.084)	(0.081)	(0.069)	(0.070)	(0.091)
Info & 90	-0.059	-0.174**	0.068	-0.131*	0.199**	0.143
	(0.082)	(0.087)	(0.095)	(0.072)	(0.080)	(0.099)
Democrat × Info	-0.082	-0.185	0.028	0.031	-0.004	0.171
	(0.117)	(0.112)	(0.115)	(0.097)	(0.105)	(0.144)
Democrat × Info & 90	0.095	0.282**	-0.025	0.075	-0.101	0.230
	(0.113)	(0.116)	(0.128)	(0.097)	(0.112)	(0.162)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with party affiliation. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.11. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.11. Support for specific policies by treatment status interacted with party affiliation

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Democrat	0.054	0.218	0.236	0.033	0.190	0.268*
	(0.168)	(0.165)	(0.161)	(0.146)	(0.151)	(0.145)
Prior	-0.002	-0.002	-0.002	-0.003**	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Democrat × Prior	0.004	0.001	0.003	0.006**	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Info	-0.149	-0.156	-0.037	-0.036	0.037	0.098
	(0.103)	(0.095)	(0.093)	(0.079)	(0.083)	(0.086)
Info & 90	0.194*	0.084	-0.074	-0.085	-0.156*	-0.152*
	(0.115)	(0.107)	(0.106)	(0.081)	(0.085)	(0.091)
Democrat × Info	0.052	0.120	-0.088	0.059	0.103	-0.067
	(0.138)	(0.135)	(0.136)	(0.119)	(0.119)	(0.115)
Democrat × Info & 90	-0.111	-0.103	0.137	-0.031	0.096	0.161
	(0.155)	(0.146)	(0.146)	(0.121)	(0.114)	(0.116)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1 interacted with party affiliation. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.12. Policy support by *Control* groups

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Prior	-0.002	-0.003	0.001	-0.001	0.002	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Control & 90	-0.054	-0.115	-0.094	-0.131	0.037	-0.049
	(0.098)	(0.099)	(0.085)	(0.080)	(0.079)	(0.106)
Observations	423	423	423	423	423	423
Control mean	4.02	4.24	-	-	-	2.72
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from a linear regression model. All regressions include demographic controls, and outcomes are normalized relative to the *Control* & *60* group. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* & *60* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* & *60* group means for the underlying specific policies are reported in Table 1.B.13. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.13. Support for specific policies by *Control* groups

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Prior	0.002	0.001	0.000	-0.002	-0.001	0.000
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Control & 90	-0.114	0.000	-0.169	-0.175*	-0.128	-0.091
	(0.104)	(0.100)	(0.108)	(0.094)	(0.097)	(0.102)
Observations	423	423	423	423	423	423
Control mean	4.70	4.39	4.51	3.75	3.96	4.39
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from a linear regression model. All regressions include demographic controls, and outcomes are normalized relative to the *Control* & *60* group. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* & *60* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

1.B.2.2 Treatment effects by prior

Table 1.B.14. Support for specific policies by treatment status split by prior

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Panel A: Prior $\leq$ 60						
Prior	-0.000	-0.000	-0.003	0.001	0.004	-0.001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Info	-0.171*	-0.056	-0.121	-0.077	-0.005	0.037
	(0.091)	(0.085)	(0.086)	(0.074)	(0.076)	(0.076)
Info & 90	0.136	-0.062	-0.001	-0.094	-0.055	-0.024
	(0.102)	(0.093)	(0.095)	(0.075)	(0.074)	(0.074)
Observations	835	835	835	835	835	835
Control mean	4.69	4.40	4.46	3.67	3.90	4.35
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Prior > 60						
Prior	-0.012	-0.006	-0.005	-0.003	-0.008	-0.005
	(0.009)	(0.008)	(0.007)	(0.006)	(0.005)	(0.007)
Info	-0.052	-0.223*	0.000	0.107	0.205**	0.128
	(0.128)	(0.122)	(0.119)	(0.105)	(0.103)	(0.107)
Info & 90	0.171	0.207	-0.037	-0.145	-0.247**	-0.220*
	(0.143)	(0.140)	(0.133)	(0.110)	(0.106)	(0.119)
Observations	439	439	439	439	439	439
Control mean	4.61	4.34	4.40	3.40	3.71	4.28
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.1. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Panel A reports results for subjects who estimate the percentage of the gender wage gap explained by children to be less than or equal to 60, whereas Panel B reports results for those estimating it to be above 60. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.15. Policy support by belief updating split by prior (IV estimates)

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Panel A: Prior $\leq$ 60						
Posterior-Prior	0.028**	-0.004	-0.018	-0.013	-0.005	0.087***
	(0.014)	(0.014)	(0.014)	(0.012)	(0.013)	(0.018)
Observations	835	835	835	835	835	835
Control mean	3.99	4.21	-	-	-	2.65
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Prior > 60						
Posterior-Prior	-0.057	-0.054	0.038	-0.066**	0.105***	0.040
	(0.035)	(0.037)	(0.039)	(0.030)	(0.031)	(0.044)
Observations	439	439	439	439	439	439
Control mean	3.90	4.01	-	-	-	2.77
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from instrumental variable specifications. The difference between posterior and prior beliefs about the percentage of the gender wage gap explained by children is instrumented with treatment assignment. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Table 1.B.16 reports the corresponding first-stage estimates. Panel A reports results for subjects who estimate the percentage of the gender wage gap explained by children to be less than or equal to 60, whereas Panel B reports results for those estimating it to be above 60. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.14. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.16. Belief updating by treatment status split by prior

	(1)
	Posterior-Prior
Panel A: Prior $\leq$ 60	
Info & 60	2.506***
	(0.131)
Info & 90	5.240***
	(0.178)
F-stat (instrument)	525.354
Observations	835
Control mean	1.31
Demographics	Yes
Panel B: Prior $>$ 60	
Info & 60	-1.457***
	(0.125)
Info & 90	1.582***
	(0.167)
F-stat (instrument)	158.060
Observations	439
Control mean	-0.68
Demographics	Yes

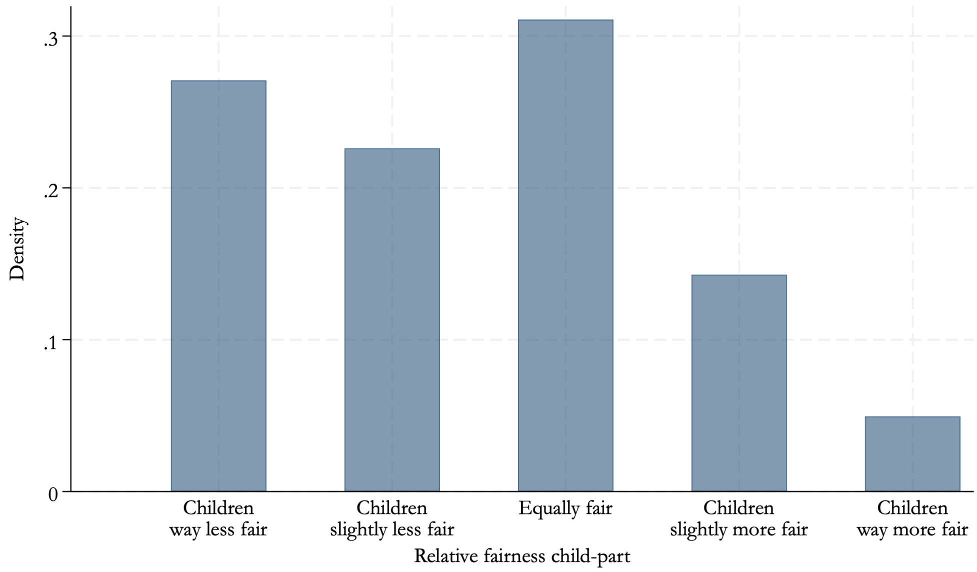
Notes: This table reports coefficient estimates from a linear regression model. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. Panel A reports results for subjects who estimate the percentage of the gender wage gap explained by children to be less than or equal to 60, whereas Panel B reports results for those estimating it to be above 60. The dependent variable is the difference between posterior and prior beliefs about the percentage of the gender wage gap explained by children *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.17. Support for specific policies by direction of information relative to prior

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Prior	0.001	0.000	-0.001	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Info	-0.137	-0.240**	0.006	0.036	0.179**	0.161*
	(0.109)	(0.109)	(0.103)	(0.091)	(0.087)	(0.090)
Info > Prior	0.125	0.193*	-0.098	-0.116	-0.204**	-0.159*
	(0.112)	(0.111)	(0.104)	(0.090)	(0.085)	(0.091)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. The dependent variables are support for maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

1.B.2.3 Policy support and fairness perceptions



Notes: This figure shows the distribution of responses to the question of whether pay differences between women and men are considered more or less fair when caused by having children, compared to other causes. Responses are measured on a 5-point Likert scale, where 1 indicates “Way less fair when caused by children" and 5 indicates “Way more fair when caused by children".

Figure 1.B.2. Perceptions about relative fairness of child-related part

Table 1.B.18. Relative assessment of child-related part by treatment groups

	Treatment group		
	Control	Info & 60	Info & 90
Rel. fairness	2.522	2.483	2.415
	(1.107)	(1.164)	(1.245)
Rel. efficiency	2.610	2.558	2.576
	(1.047)	(1.113)	(1.104)
Observations	423	441	410

Notes: This table reports mean perceptions, by treatment status, of how fair and efficient the child-related component of the gender wage gap is compared to the component unrelated to children. Responses are measured on a 5-point Likert scale, where 1 indicates “Way less fair/efficient when caused by children" and 5 indicates “Way more fair/efficient when caused by children".

Table 1.B.19. Policy support by direction of information relative to prior interacted with fairness perceptions

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child- Discrim.	(6) Priorit. child-part
Prior	-0.006***	-0.006**	-0.003	-0.004**	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child less fair	-0.101	-0.005	0.012	0.053	-0.041	-0.139
	(0.170)	(0.164)	(0.150)	(0.140)	(0.146)	(0.187)
Child less fair × Prior	0.008***	0.007**	0.006**	0.006**	0.000	0.006*
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	0.355***	0.292**	-0.122	0.234**	-0.356***	0.085
	(0.133)	(0.130)	(0.140)	(0.107)	(0.119)	(0.135)
Child less fair × Info	-0.215	-0.378**	-0.009	-0.244*	0.235	0.022
	(0.167)	(0.179)	(0.165)	(0.139)	(0.149)	(0.215)
Info > Prior	-0.361***	-0.373***	-0.062	-0.307***	0.245**	0.022
	(0.136)	(0.134)	(0.148)	(0.110)	(0.122)	(0.141)
Child less fair × Info > Prior	0.390**	0.478***	0.224	0.274*	-0.050	0.287
	(0.166)	(0.179)	(0.172)	(0.140)	(0.151)	(0.218)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less fair than the part unrelated to children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variables *Info=Prior* and *Child less fair×Info=Prior* are included in all specifications but omitted from the table due to the small number of observations. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.4. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.20. Support for specific policies by direction of information relative to prior split by fairness perceptions

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Panel A: Equally or more fair						
Prior	-0.002	-0.002	-0.003	-0.005**	-0.002	-0.003
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Info	-0.172	-0.209	0.040	0.237*	0.280**	0.208
	(0.179)	(0.159)	(0.154)	(0.122)	(0.119)	(0.141)
Info > Prior	0.032	0.081	-0.309*	-0.346***	-0.337***	-0.233
	(0.189)	(0.166)	(0.161)	(0.124)	(0.121)	(0.144)
Observations	641	641	641	641	641	641
Control mean	4.55	4.24	4.31	3.30	3.59	4.15
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Less fair						
Prior	0.004***	0.004**	0.002	0.003	0.002	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Info	-0.099	-0.303**	-0.027	-0.203	0.052	0.123
	(0.087)	(0.126)	(0.130)	(0.132)	(0.121)	(0.093)
Info > Prior	0.158*	0.287**	0.077	0.118	-0.070	-0.120
	(0.088)	(0.125)	(0.125)	(0.130)	(0.116)	(0.092)
Observations	633	633	633	633	633	633
Control mean	4.80	4.55	4.59	3.90	4.11	4.54
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

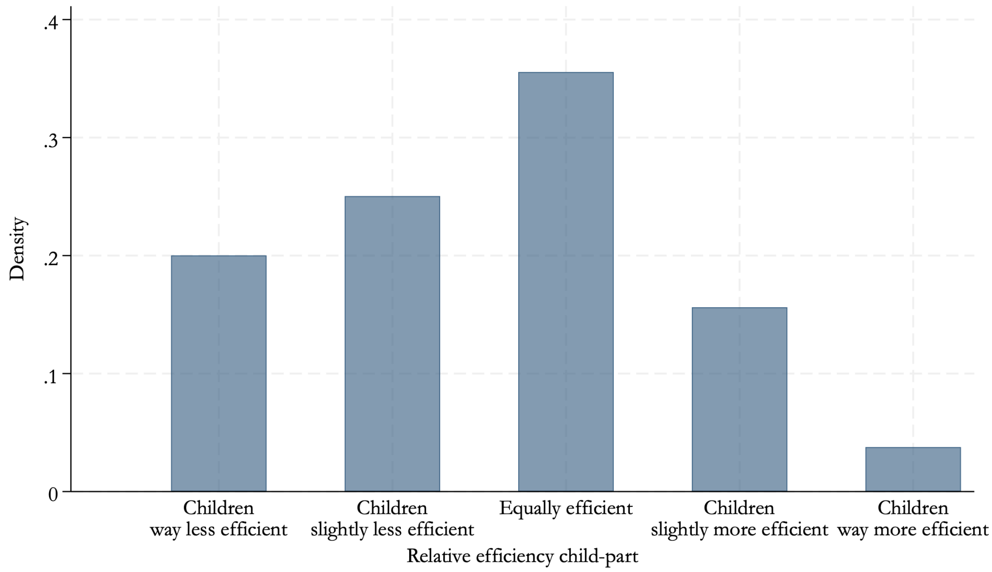
Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. Panel A reports results for respondents who believe that the child-related part is equally or more fair, whereas Panel B reports results for those who believe it is less fair than the part not explained by children. The dependent variables are support for the following policies: maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.21. Support for specific policies by direction of information relative to prior interacted with fairness perceptions

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Prior	-0.002	-0.003	-0.003	-0.006***	-0.003	-0.004*
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child less fair	0.013	-0.016	0.038	-0.074	0.122	0.110
	(0.182)	(0.178)	(0.173)	(0.160)	(0.165)	(0.174)
Child less fair × Prior	0.006**	0.007**	0.004	0.008***	0.004	0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Info	-0.190	-0.206	0.030	0.229*	0.273**	0.200
	(0.177)	(0.161)	(0.153)	(0.122)	(0.119)	(0.141)
Child less fair × Info	0.112	-0.077	-0.063	-0.429**	-0.215	-0.089
	(0.199)	(0.208)	(0.198)	(0.179)	(0.170)	(0.169)
Info > Prior	0.039	0.081	-0.305*	-0.342***	-0.337***	-0.240*
	(0.186)	(0.169)	(0.160)	(0.124)	(0.121)	(0.144)
Child less fair × Info > Prior	0.112	0.183	0.377*	0.448**	0.250	0.125
	(0.207)	(0.212)	(0.202)	(0.179)	(0.168)	(0.171)
Observations	1274	1274	1274	1274	1274	1274
Control mean	4.66	4.38	4.44	3.58	3.83	4.33
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less fair than the part unrelated to children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variables *Info=Prior* and *Child less fair×Info=Prior* are included in all specifications but omitted from the table due to the small number of observations. The dependent variables are support for the following policies: maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

1.B.2.4 Policy support and efficiency perceptions



Notes: This figure shows the distribution of responses to the question of whether pay differences between women and men are considered more or less efficient when caused by having children, compared to other causes. Responses are measured on a 5-point Likert scale, where 1 indicates "Way less efficient when caused by children" and 5 indicates "Way more efficient when caused by children".

Figure 1.B.3. Perceptions about relative efficiency of child-related part

Table 1.B.22. Policy support by direction of information relative to prior split by efficiency perceptions

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Panel A: Equally or more efficient						
Prior	-0.004**	-0.004**	-0.001	-0.002	0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Info	0.183	0.034	-0.207	0.054	-0.261**	0.102
	(0.131)	(0.130)	(0.129)	(0.105)	(0.115)	(0.135)
Info > Prior	-0.136	-0.173	0.076	-0.117	0.192	0.013
	(0.134)	(0.131)	(0.133)	(0.107)	(0.117)	(0.141)
Observations	700	700	700	700	700	700
Control mean	3.82	4.05	-	-	-	2.67
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Less efficient						
Prior	0.000	-0.000	0.001	-0.001	0.002	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Info	0.292***	0.164	-0.050	0.132	-0.182*	0.057
	(0.109)	(0.131)	(0.110)	(0.096)	(0.100)	(0.180)
Info > Prior	-0.166*	-0.068	0.051	-0.189**	0.240**	0.372**
	(0.097)	(0.125)	(0.116)	(0.092)	(0.102)	(0.182)
Observations	574	574	574	574	574	574
Control mean	4.17	4.28	-	-	-	2.72
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. Panel A reports results for respondents who believe that the child-related part is equally or more efficient, whereas Panel B reports results for those who believe it is less efficient than the part not explained by children. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.23. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.23. Support for specific policies by direction of information relative to prior split by efficiency perceptions

	(1) Matern. leave	(2) Patern. leave	(3) Child care	(4) Quota	(5) AAP	(6) Equal pay
Panel A: Equally or more efficient						
Prior	-0.001	-0.000	-0.001	-0.002	-0.001	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Info	-0.271*	-0.301*	-0.049	-0.010	0.107	0.066
	(0.164)	(0.157)	(0.145)	(0.119)	(0.124)	(0.136)
Info > Prior	0.177	0.205	-0.156	-0.037	-0.162	-0.151
	(0.167)	(0.160)	(0.148)	(0.121)	(0.124)	(0.140)
Observations	700	700	700	700	700	700
Control mean	4.59	4.28	4.33	3.42	3.63	4.21
Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Less efficient						
Prior	0.003	0.002	-0.000	-0.002	-0.000	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Info	-0.005	-0.199	0.054	0.010	0.158	0.227**
	(0.121)	(0.141)	(0.138)	(0.141)	(0.121)	(0.103)
Info > Prior	0.035	0.175	-0.056	-0.174	-0.215*	-0.178*
	(0.130)	(0.149)	(0.136)	(0.137)	(0.115)	(0.097)
Observations	574	574	574	574	574	574
Control mean	4.76	4.53	4.59	3.80	4.12	4.50
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variable *Info=Prior* is included in all specifications but omitted from the table due to the small number of observations. Panel A reports results for respondents who believe that the child-related part is equally or more efficient, whereas Panel B reports results for those who believe it is less efficient than the part not explained by children. The dependent variables are support for the following policies: maternity leave (column (1)), paternity leave (column (2)), child care subsidies (column (3)), gender quotas (column (4)), affirmative action (column (5)), and equal pay legislation (column (6)). *Control* group means prior to normalization are reported at the bottom of the table and are calculated within the respective subsample. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.24. Policy support by direction of information relative to prior interacted with efficiency perceptions

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Prior	-0.004**	-0.004**	-0.001	-0.002	0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child less efficient	0.039	-0.083	0.083	0.179	-0.095	-0.363*
	(0.165)	(0.161)	(0.150)	(0.137)	(0.140)	(0.195)
Child less efficient × Prior	0.004	0.004	0.002	0.001	0.001	0.008**
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Info	0.184	0.041	-0.221*	0.065	-0.285**	0.098
	(0.129)	(0.129)	(0.130)	(0.105)	(0.115)	(0.134)
Child less efficient × Info	0.128	0.158	0.187	0.105	0.082	-0.028
	(0.169)	(0.183)	(0.168)	(0.142)	(0.152)	(0.221)
Info > Prior	-0.133	-0.176	0.088	-0.123	0.211*	0.016
	(0.133)	(0.130)	(0.134)	(0.106)	(0.117)	(0.140)
Child less efficient × Info > Prior	-0.041	0.099	-0.039	-0.084	0.045	0.370
	(0.165)	(0.180)	(0.174)	(0.141)	(0.154)	(0.226)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less efficient than the part unrelated to children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variables *Info=Prior* and *Child less efficient×Info=Prior* are included in all specifications but omitted from the table due to the small number of observations. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.4. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

1.B.2.5 Beliefs about sources of the gender wage gap

Table 1.B.25. Beliefs about sources by direction of information relative to prior interacted with fairness perceptions

	(1) Individ. choice	(2) Discrim.	(3) Prefer.	(4) Caregiv. skills	(5) Job-rel. skills	(6) Work perform.
Prior	0.005**	-0.006***	0.002	0.001	0.002	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Child less fair	0.140	-0.186	0.063	-0.078	0.041	-0.005
	(0.177)	(0.173)	(0.178)	(0.164)	(0.167)	(0.157)
Child less fair × Prior	-0.007**	0.008***	-0.004	-0.002	-0.002	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Info	-0.275**	0.199	-0.372***	-0.136	-0.265**	-0.052
	(0.125)	(0.125)	(0.123)	(0.129)	(0.117)	(0.121)
Child less fair × Info	0.288	-0.330*	0.275	0.150	0.201	0.022
	(0.191)	(0.182)	(0.187)	(0.191)	(0.173)	(0.176)
Info > Prior	0.296**	-0.244*	0.277**	0.061	0.199	0.110
	(0.125)	(0.133)	(0.123)	(0.130)	(0.122)	(0.119)
Child less fair × Info > Prior	-0.389**	0.428**	-0.272	-0.080	-0.219	-0.151
	(0.187)	(0.182)	(0.183)	(0.190)	(0.173)	(0.171)
Observations	1274	1274	1274	1274	1274	1274
Control mean	2.65	3.32	3.08	2.80	3.03	2.58
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less fair than the part unrelated to children. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variables *Info=Prior* and *Child less fair×Info=Prior* are included in all specifications but omitted from the table due to the small number of observations. The table presents results for beliefs about the degree to which different sources of inequality are driving the gender wage gap. Sources include individual choices (column (1)), discrimination (column (2)), differences in preferences (column (3)), caregiving skills (column (4)), job-related skills (column (5)) and work performance (column (6)). Robust standard errors are displayed in parentheses. *Control* group means prior to normalization are reported at the bottom of the table. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Table 1.B.26. Policy support by direction of information relative to prior interacted with fairness perceptions and controlling for beliefs about sources

	Overall support		Specific support		Relative support	
	(1) General support	(2) Govern. interv.	(3) Child policies	(4) Discrim. policies	(5) Child-Discrim.	(6) Priorit. child-part
Prior	-0.003*	-0.002	-0.001	-0.001	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Child less fair	0.001	0.097	0.073	0.140	-0.066	-0.130
	(0.133)	(0.131)	(0.144)	(0.115)	(0.144)	(0.187)
Child less fair × Prior	0.004	0.003	0.003	0.002	0.001	0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Info	0.196*	0.121	-0.210	0.108	-0.318***	0.074
	(0.111)	(0.109)	(0.137)	(0.093)	(0.116)	(0.137)
Child less fair × Info	-0.008	-0.166	0.103	-0.080	0.182	0.033
	(0.147)	(0.166)	(0.165)	(0.127)	(0.146)	(0.215)
Info > Prior	-0.190*	-0.200*	0.038	-0.174*	0.212*	0.046
	(0.110)	(0.112)	(0.144)	(0.097)	(0.120)	(0.143)
Child less fair × Info > Prior	0.127	0.221	0.074	0.071	0.003	0.253
	(0.144)	(0.163)	(0.170)	(0.127)	(0.149)	(0.217)
Individual choice	-0.089***	-0.070***	-0.086***	-0.071***	-0.015	-0.150***
	(0.025)	(0.026)	(0.027)	(0.022)	(0.028)	(0.039)
Discrimination	0.447***	0.433***	0.229***	0.360***	-0.131***	-0.040
	(0.026)	(0.027)	(0.027)	(0.023)	(0.026)	(0.036)
Preferences	-0.084***	-0.119***	-0.036	-0.083***	0.047	0.049
	(0.026)	(0.027)	(0.028)	(0.023)	(0.030)	(0.041)
Caregiving skills	-0.011	0.011	0.045	0.037*	0.008	0.003
	(0.027)	(0.028)	(0.030)	(0.022)	(0.029)	(0.040)
Job-related skills	-0.045	-0.096***	-0.040	-0.044	0.005	0.031
	(0.035)	(0.034)	(0.035)	(0.029)	(0.035)	(0.048)
Work performance	-0.025	0.050	-0.026	0.053*	-0.078**	-0.087*
	(0.033)	(0.034)	(0.039)	(0.029)	(0.038)	(0.046)
Observations	1274	1274	1274	1274	1274	1274
Control mean	3.96	4.14	-	-	-	2.69
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports coefficient estimates from linear regressions estimating equation 1.2 interacted with whether the participant perceives the child-related part of the gender wage gap as less fair than the part unrelated to children, and controlling for beliefs about its sources. All regressions include demographic controls, and outcomes are normalized relative to the *Control* group. The dummy variables *Info=Prior* and *Child less fair×Info=Prior* are included in all specifications but omitted from the table due to the small number of observations. Columns (1)–(2) present results for overall support, distinguishing between general support for reducing the gap (column (1)) and support for government intervention to achieve this (column (2)), columns (3)–(4) for specific policies, and columns (5)–(6) for child-related policies relative to other policies. Column (3) reports results for support for child-related policies, measured as the average of normalized support for maternity leave, paternity leave, and child care subsidies. Column (4) reports results for support for discrimination-related policies, measured as the average of normalized outcomes for support for quotas, affirmative action plans, and equal pay legislation. Column (5) shows the difference between support for child-related and discrimination-related policies, while column (6) captures respondents' preference for prioritizing pay differences related to having children. *Control* group means prior to normalization are reported at the bottom of the table. For outcomes constructed from multiple variables, the *Control* group means for the underlying specific policies are reported in Table 1.B.4. Robust standard errors are displayed in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Chapter 2

Age at Labor Market Entry and Gender Occupational Segregation*

Joint with Anna Person

2.1 Introduction

Gender differences in occupational sorting remain a central driver of persistent wage inequality (Goldin and Katz, 2016; Goldin et al., 2017). Despite extensive research on preferences, norms, and discrimination, relatively little is known about how the timing of labor market entry shapes gendered career trajectories (see e.g., Penner, 2008; Gartner and Hinz, 2009). This is surprising, as the transition from school to work is a critical juncture in which occupational paths often become persistent.

This paper studies whether age at labor market entry causally affects gender differences in occupational choice. We focus on timing as a policy-relevant and largely overlooked mechanism: entering the labor market at a younger age may amplify gendered peer dynamics, social identity concerns, or stereotypical beliefs at the moment of choice. In this way, age can shape whether young workers sort into male- or female-dominated occupations. We find that younger men are significantly more likely to select into male-dominated occupations, whereas women's occupational choices appear unaffected by age at entry.

* We are especially grateful to Radost Holler, whose initiative was fundamental in the early stages of this project. We also thank Klaus Altmos, Anna Bindler, Barbara Boelmann, Teodora Boneva, Thomas Dohmen, Antonia Entorf, Christina Felfe, Daniel Kühnle, Valentina Melentyeva, Pia Pinger, Jan Scherer, Malin Siemers, and Jakob Wegmann for their valuable comments and support. We are also thankful to the participants at the ECONtribute Junior Workshop in Gender and Labor Economics, the Essen Health Conference, the Bonn Applied Micro Coffee, the RGS Doctoral Conference, and the NGE Workshop for helpful feedback and suggestions. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866 and through the Collaborative Research Center Transregio 224 (CRC TR 224 - project A02).

Age at labor market entry is particularly relevant given that research has shown age at important life junctures can have lasting effects, especially in youth.[1] Furthermore, occupational choice often occurs at the end of formal education and is frequently a one-time decision. Given that half of all workers remain in their initial occupation long-term (SOEP, own calculations), the age at choosing could significantly impact career trajectories. This is particularly relevant for lower-educated individuals who enter the labor market earlier. Since younger individuals have less information on the labor market, they may, by default, turn to more stereotypical occupations due to these skewed beliefs.

This phenomenon can be observed very well in countries with tracking systems of education that exhibit different durations of secondary education. Thus, Germany provides an ideal setting for studying the effect of age at labor market entry. Germany has an early-tracking Vocational Education and Training (VET) system after compulsory education, forcing children to make an occupational choice at a relatively young age. In this context, if age indeed influences occupational choice, even a difference of one year could have a significant impact on an individual's career trajectory. Moreover, age could have a differential impact on occupational choice by gender, especially considering that men in Germany are less likely to enroll in university and thus face earlier occupational decisions. As a consequence, gender occupational segregation is more pronounced among lower-educated individuals: one in two lower-educated men works in a male-segregated occupation,[2] compared to only one in six women working in a female-segregated one (German Microcensus, own calculations). Among the higher-educated, one in five men and one in thirteen women work in a gender-segregated occupation aligned with their own gender.

Providing causal estimates of the effect of age on labor market entry decisions, i.e., occupational choice, can be difficult because of selection and school class effects. Usually, a class is uniformly affected by events that occur at specific points in time, and within-class differences in age are driven by selection bias, causing age to correlate with unobserved characteristics. To overcome this, we exploit a series of reforms in Germany that shifted school entry cutoff dates in some states between 2003 and 2010, thus generating exogenous variation in effective school entry (and exit) for children born between 1997 and 2004.

Using a Regression Discontinuity Design (RDD) with school entry dates as a cutoff, we find that men are on average 7% more likely to choose a male-segregated

1. For example, being the oldest in class is associated with higher confidence (Page, Sarkar, and Silva-Goncalves, 2019), better performance in competitive elite sports (Allen and Barnsley, 1993), and increased likelihood of attaining managerial and political positions (Du, Gao, and Levi, 2012; Muller and Page, 2016). These individuals also show higher self-esteem (Thompson, Barnsley, and Battle, 2004), fewer behavioral problems (Chen, Fortin, and Phipps, 2015), and more competitiveness (Page, Sarkar, and Silva-Goncalves, 2017).

2. Gender-segregated occupations are defined here as occupations where less than 10% are other-gender workers.

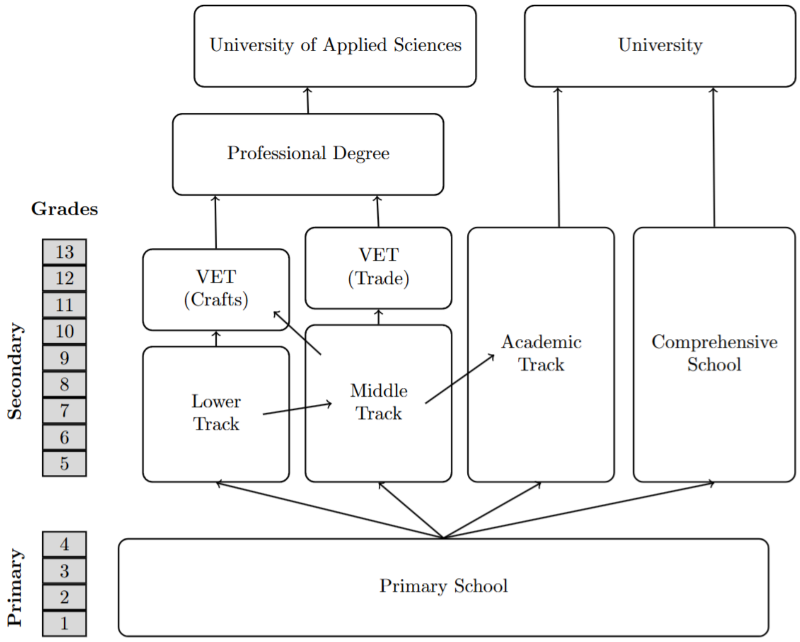
occupation when starting (and leaving) school one year earlier. We do not find a significant impact of age at labor market entry on occupational segregation for women. Using detailed individual-level data, we can show that this is not driven by differences in educational attainment, track choice, or relative age within the class. Instead, intergenerational transmission of parental occupations, differential educational returns, and the onset of stereotypes are potential drivers of the effect.

Our paper makes several contributions. First, by identifying a key factor of occupational choice, namely age at labor market entry, we contribute to the literature on the determinants of educational and occupational choices (Levine, 1976; Keane and Wolpin, 1997; Attanasio and Kaufmann, 2017) by presenting a new and policy-relevant channel. Existing studies on occupational choice primarily focus on two areas: personal beliefs or preferences regarding job characteristics (Antecol and Cobb-Clark, 2013; Zafar, 2013; Cortes and Pan, 2018; Wiswall and Zafar, 2018; Vattuone, 2026) and intergenerational transmission of occupations (Vleuten et al., 2018; Humlum, Nandrup, and Smith, 2019; Philipp, 2023). Therefore, we are adding an important determinant related to the one-off, time sensitive nature of occupational choice. Second, our research contributes to the literature investigating the persistence of gender segregation in the labor market. In particular, we identify a relevant factor that could explain men's attachment to traditionally male-segregated occupations (Delfino, 2024). Specifically, the problem could potentially be less severe if this decision is made later in life. Third, by focusing on lower-educated individuals, we also extend the literature on early tracking and inequality in Germany (Gamoran and Mare, 1989; Slavin, 1990; Oakes, 2005). Since gender occupational segregation is more persistent and pronounced among nonacademic tracks, we contribute to the debate on equity concerns of VET and early occupational sorting. Specifically, concerns that early tracking may constrain individuals' future educational and occupational opportunities by requiring career-relevant decisions at a very young age.

The remainder of the paper is structured as follows: Section 2.2 provides an overview of the German education system and the relevant reforms. It further introduces the data sources. Section 2.3 presents the empirical strategy and discusses our main results. Section 2.4 explores potential mechanisms and provides robustness checks of our results. Finally, section 2.5 concludes by discussing policy implications and suggestions for future research.

2.2 Institutional background and data

In Germany, education falls under the authority of the federal states. To coordinate certain standards, however, the assembly of ministers of education of the German states, "Kultusministerkonferenz" (KMK), regularly decides on non-binding directives for harmonization. Thus, while state-level differences exist, the core structure of the school system is largely consistent across the country.[3]



Notes: This figure shows the structure of the German education system. Source: Authors' own illustration.

Figure 2.1. Structure of the German education system

Children traditionally enter primary school at age six or seven depending on their date of birth. Primary school lasts four to six years depending on the state.[4] After primary school, children are sorted into different secondary school tracks, each having different lengths and content, and yielding different school leaving certificates. There are four main secondary school tracks:

(1) **Academic track** (Gymnasium) lasting eight or nine and yielding a university entrance qualification[5]
(2) **Middle track** (Realschule) lasting for six years and qualifying for all VET programs

3. It is important to note that the states that were part of the GDR displayed a markedly different system between 1946 and 1990. After the reunification with the school year 1991/92, the school system of the states of the former GDR was harmonized towards the system of the West German states (see e.g., Fuchs and Reuter, 1995, for details). In Saxony, the harmonization was delayed to the school year 1992/93 (Fuchs and Reuter, 1995).

4. Berlin and Brandenburg are the only states where the length of primary school is six grades. In all other states, it is four years long.

5. Whether it is eight or nine years depends on the state as well as the cohort. Originally it was nine years in all West German states and eight in the East German states. However, in the 2000's there was a movement to decrease its length to eight years also in the West German states. This was later partially reverted again so that it truly depends on the cohort whether individuals did eight or nine years of 'Gymnasium'.

(3) **Lower track** (Hauptschule) lasting for five to six years and qualifying for a subset of VET programs [6]

(4) **Comprehensive school** (e.g., Gemeinschaftsschule, Gesamtschule, Sekundarschule) combining different tracks in one school

The choice of school track is based entirely on the recommendation of the primary school teacher in the states where it is binding (Baden-Württemberg, Bavaria, Berlin, Brandenburg and Thuringia). In the other states, children also receive a recommendation by the teacher, but the choice is eventually made by the parents. Transition across tracks is possible, however its structure and pervasiveness depends on the state and on the track choice. Most of the transitions from the lower tracks to the higher tracks happen after graduation from the respective lower secondary school. On average, across the whole secondary school period, slightly more than one student out of four changes their initial track (Bildungsberichterstattung, 2020).

In principle, schooling is mandatory until the child reaches eighteen years of age in most states. However, full-time general education is only mandatory for nine years (ten years in Brandenburg, North Rhine-Westphalia, and Berlin). After completing lower secondary education, students have to either continue with full-time education in any other type of upper secondary school or complete their compulsory education by participating in a VET program with at least part-time schooling (also known as the 'dual system'). This school-to-work transition stage is central for our setting because many occupational choices in Germany are made at that point.

There are two types of VET: school-based or on-the-job training with part-time schooling called the 'dual system'. The type of training depends on the occupations students want to be trained for. School-based training is mostly for social, health and medical professions, such as physiotherapy, speech and language therapy, medical and technical assistant, or similar. Most occupations that do not require university education require training within the 'dual system' (330 occupations are trained within the dual system). In the 'dual system', students are employed and trained on the job. They typically visit school only in certain weeks during the year. After completion, they obtain a formal vocational degree, which is often required to practice legally protected occupations. This makes the apprenticeship track highly relevant for our analysis, as it is the dominant route into many occupations where gender segregation is pronounced.

2.2.1 School entry age in Germany

In this paper, we mainly rely on one particular feature of the German educational system, namely a discontinuity in the school starting age. Each German state has a certain school starting cutoff date. Every child aged six before that cutoff date is

6. Six years only in North-Rhine Westphalia.

meant to enter primary school in August or September of that year, while every child born after that cutoff date is meant to enter primary school a year later.

Historically, the cutoff date has been June 30 every year (see Appendix B for details on the evolution of the cutoff dates across states).[7] That is, children born before June 30 enter school in the year they turn six, while those born after June 30 enter school in the year they turn seven.

In 1997, the KMK agreed to allow more flexibility in school-entry cutoff dates, with the aim of encouraging earlier school entry. Specifically, states could move the cutoff from June 30 to a date up to September 30 (see e.g., Faust, 2006). In response, eight of the sixteen federal states changed their month-of-birth cutoff between 2003 and 2010.

For our design, the key point is not the reform process itself, but the exogenous timing variation in school-entry age generated by these cutoff rules. Importantly, the cutoff changes occurred after the relevant children had already been born, which limits concerns that parents could sort births strategically around the new thresholds. Table 2.1 reports the affected states, implementation years, and first affected cohorts, and Figure 2.B.1 illustrates the geographic pattern.[8]

It is important to note that cutoffs are not fully binding. Both earlier and delayed entry are, in principle, possible if certain conditions are met. Children can enter early if the child displays the physical, intellectual and social development ('Schulfähigkeit') to enter school. Most states do not stipulate any further restriction on early entry. In some states, however, early entry is only permitted for children who are not yet six years old at the normal cutoff date, but turn six within a certain time frame after the cutoff. The time frame depends on the state. In most states, a delay of one year is allowed if the child does not display the characteristics needed to enter school according to a pediatrician's assessment ('Schulfähigkeitstest', school readiness test, Horstschräer and Muehler, 2014). We provide additional information on the institutional background in Appendix 2.B.

2.2.2 Data

We use the German Microcensus to identify the effect of interest in a German representative sample with a large number of observations. Provided by the Research Data Centres of the Statistical Offices of the Federation and the Federal States (RDC), the German Microcensus is a multiple-subject survey using a representative sample of one percent of the German population. It is carried out annually since 1957 and includes about 370 000 households with 810 000 household members. The Microcen-

7. To be precise, in the German Democratic Republic the cutoff was May 31. After reunification, East German states moved their cutoff subsequently to June 30. For more details see Appendix B.

8. In Berlin, Bavaria, and Lower Saxony, cutoff dates were later shifted back in the 2010s and 2020s. These reversals do not affect the cohorts observed in our sample and are therefore not used in our analysis.

Table 2.1. Reforms across federal states

Federal State *- Year of Reform*	**Cutoff Date before Reform**	**Cutoff Date after Reform**	**Pivotal Cohort**	**Entry in Education**
Thuringia				
- 2003	June 30	Aug 1	1997	2003
Baden-Württemberg				
- 2005	June 30	July 31	1999	2005
- 2006	July 31	Aug 31	2000	2006
- 2007	Aug 31	Sept 30	2001	2007
Bavaria				
- 2005	June 30	July 31	1999	2005
- 2006	July 31	Aug 31	2000	2006
- 2007	Aug 31	Sept 30	2001	2007
- 2008	Sept 30	Oct 31	2002	2008
- 2009	Oct 31	Nov 30	2003	2009
- 2010	Nov 30	Sept 30	2004	2010
Berlin				
- 2005	June 30	Dec 31	1999	2005
- 2017	Dec 31	Sept 30	2011	2017
- 2022	Sept 30	June 30	2016	2022
Brandenburg				
- 2005	June 30	Sept 30	1999	2005
North Rhine-Westphalia				
- 2007	June 30	July 31	2001	2007
- 2008	July 31	Aug 31	2002	2008
- 2009	Aug 31	Sept 30	2003	2009
Rhineland-Palatinate				
- 2008	June 30	Aug 31	2002	2008
Lower Saxony				
- 2010	June 30	July 31	2004	2010
- 2011	July 31	Aug 31	2005	2011
- 2012	Aug 31	Sept 30	2006	2012
- 2018	Sept 30	June 30	2012	2018

Notes: The table gives an overview of when the reforms were implemented in the eight federal states that changed their cutoff dates from June 30. It is ordered by the year of implementation of the reform that coincides with the entry in education for the pivotal cohort. The cutoff dates before and after the reform are also displayed. The default cutoff for all other federal states is still June 30. Sources: Faust (2006), Berthold (2008), Hövel (2009), and Helbig and Nikolai (2015) sus has detailed information on labor market outcomes, such as occupational choice, educational trajectories, as well as state, year and month of birth. We have obtained six survey years from the RDC for On-Site use: 2014, 2015, 2016, 2017, 2018 and 2019.[9] In these years, we can access information on the labor market choices of the pivotal cohorts entering the labor market in the years 2015-2018. We can analyze all states that introduced the reforms in the years 2003 up to 2010. The cross-sectional structure of the data allows us to pool information on the pivotal cohorts and increase the power of our analysis. Our realized sample consists of 29,598 individuals for which we observe labor market outcomes already.

One drawback of this data source is that it does not contain information on the year individuals started school. To correct for this, we make use of the German Socio-Economic Panel (SOEP), a longitudinal study of private households in Germany, conducted annually since 1984 (DIW, 2022). It provides representative microdata on individuals and households, covering a wide range of topics including employment, education, income, and demographic characteristics. Although the SOEP's long-run panel is ideal to analyze detailed educational trajectories of young individuals, its small sample size relative to our relevant cohorts (n = 3,088)[10] is not enough to estimate our main effect of interest.

We employ both datasets for complementary purposes in our empirical strategy. The SOEP serves as our primary data source for validating the relevance assumption of our fuzzy regression discontinuity design, providing direct evidence of the first-stage relationship between cutoff dates and school entry timing (Table 2.2, column 1), as well as confirming the persistence of this effect on age at school exit (Table 2.3). The Microcensus constitutes our main analytical dataset for estimating the treatment effects on occupational segregation (Table 2.6), leveraging its large sample size and comprehensive occupational information. Additionally, we use the Microcensus to construct our segregation measures based on the full distribution of workers across occupations (Figure 2.3) and to provide supplementary first-stage evidence for younger cohorts where direct school entry timing is observable (Table 2.2, columns 2-6).

2.3 Empirical strategy and results

2.3.1 Empirical strategy

This paper examines the causal relationship between age at labor market entry and gender occupational segregation. Our identification strategy leverages exogenous

9. RDC of the Federal Statistical Office and Statistical Offices of the Federal States of Germany, Microcensus, survey year(s) 2014-2015.

10. These are the observations of our relevant cohorts for which we have information on age at school entry in the SOEP. The number of observations reduces to 75 if we consider individuals affected by the reforms for which we have nonmissing occupational info.

variation in school entry age induced by state-level cutoff dates reforms in Germany. We employ a fuzzy regression discontinuity design (RDD) by looking at career choices of young adults born just before and after the cutoff date; these individuals, although of very similar age, start school one year apart, thus entering the labor market at different ages.

Our preferred and most credible source of identifying variation is the discontinuity in school entry probability at the cutoff: comparing individuals born just before and just after the cutoff date within the same cohort and state, we can isolate the effect of starting school — and thus entering the labor market — at a younger age. The state-level reforms generate an additional and distinct source of variation, namely cross-cohort and cross-state differences in cutoff timing. We exploit this variation in our robustness analysis, where we implement an RDD-DiD design comparing cohorts affected by the reforms to those that were not within the same states. This serves to validate that our main RDD estimates are not driven by pre-existing differences across cohorts (see Section 2.4.2).

For the effect to be causal, we need the following assumptions to hold: (i) there is no manipulation of the assignment variable (exogeneity); (ii) the effect due to the discontinuity occurs only through the assignment variable (exclusion restriction); (iii) the instrument shifts treatment in the same direction for all individuals (monotonicity); and (iv) the cutoff determines a substantial shift in the probability of being treated (relevance).

2.3.1.1 Exogeneity and the exclusion restriction

The primary identification challenge stems from the potential endogeneity between labor market entry age and occupational choice. The most significant confounding factor is school track selection, which simultaneously influences both the duration of education and subsequent occupational choices. While track choice is endogenous in our setup, we demonstrate that the reforms we exploit do not systematically alter track selection patterns (see Section 2.4).

Even within educational tracks, school entry age creates variation in labor market entry age. This variation is partially exogenous, as it is determined by state-level cutoff dates based on children's month of birth. Two potential threats to identification exist. First, parents might strategically time births to influence their children's school entry age. However, our data show little evidence of such behavior (see 2.A.4 and 2.A.2) when we test for manipulation as proposed by Frandsen (2017).

Second, parents have some discretion in school entry timing ("redshirting"), although our data indicate high compliance rates. Specifically, 82% of children comply with the designated cutoff dates. Among non-compliers (18%), 61% (11%) enter school earlier than designated, while 39% (7%) enter school later than designated (DIW, 2022). A residual concern is that parents who send their children to school earlier regardless of the cutoff may also independently direct them toward gender-segregated occupations – for instance, out of a preference for occupations that do not require university qualifications – thereby confounding our estimates. However, evidence from Germany suggests that, in contrast to the US and other countries where higher-educated parents are more likely to redshirt their children (Black, Devereux, and Salvanes, 2011), it is children from *lower* SES households who tend to be sent to school later (Kratzmann and Schneider, 2009). Parents who send their children to school early in our setting are therefore more likely to be of higher SES – precisely those less likely to steer their children toward gender-segregated occupations. If anything, this implies that our estimates constitute a lower bound of the true effect.

To strengthen our identification strategy, we exploit reforms that shifted school eligibility rules based on month of birth cutoff dates. These reforms provide an exogenous source of variation in school entry age for several reasons. First, the reforms were implemented *after* the affected cohorts were born, preventing anticipatory responses from parents. Second, the changes in eligibility rules were determined at the state level, independent of individual characteristics. Third, the timing and magnitude of these reforms varied across states, as shown in Section 2.2, providing additional variation. This approach allows us to identify the causal effect of labor market entry age on gender occupational segregation while addressing the aforementioned endogeneity concerns.

2.3.1.2 Monotonicity

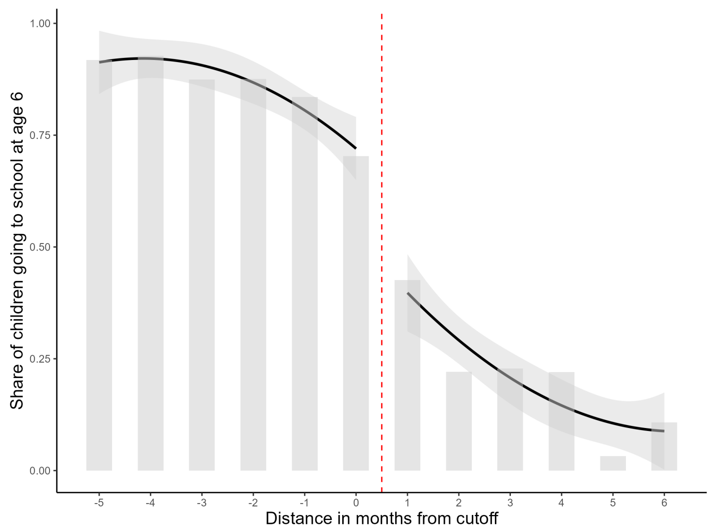
The presence of non-compliance in both directions – children entering school earlier (11% of the full sample) and later than prescribed (7%) – raises the question of whether monotonicity holds in our design. In a fuzzy RD design, monotonicity requires that the instrument shifts treatment in the same direction for everyone: no child who would have been sent to school early if born *before* the cutoff would have been sent to school late if born *after* it. Two-sided non-compliance introduces the possibility of defiers, whose presence would invalidate the LATE interpretation, as discussed by Barua and Lang (2016).

While we cannot rule out the presence of defiers from a purely formal perspective, we argue that their practical relevance is limited in our setting. For defiers to bias our estimates, their potential occupational outcomes would need to differ systematically from those of compliers in a way that correlates with the direction in which they defy the rule. It is not clear what mechanism would support such a correlation: defier status reflects a family's tendency to go against the prescribed school entry rule regardless of its direction, and it is difficult to construct a plausible story linking this behavioral pattern to preferences over the gender composition of a child's future occupation. This is especially so given that our outcome is measured nine to eleven years after school entry, making it implausible that a parental timing decision at age five exerts a meaningful independent influence on occupational choices made a decade later.

2.3.1.3 Relevance

The last assumption in our design implies (i) that school cutoff entry rules significantly shift the probability for individuals to start school in a certain year, and (ii) that age at school entry affects age at school exit, which in turn affects age at labor market entry.

As highlighted before, we rely on a fuzzy design because school entry rules are not completely binding, and parents have some leeway, i.e. the cutoff is not sharp. We begin by showing in Figures 2.2, 2.A.2 and 2.A.3 that at the cutoff date there is a substantial shift in the probability of starting school in the cohorts affected by the reforms.



Notes: The figure shows the share of children starting school the year they turn six in the cohorts from 1996 to 2004. On the x axis, month of birth is normalized as to be equal to 0 for the last month before the cutoff date, and is equal to 1 for the first month after the cutoff. That is, in the states without reforms, for example, the cutoff is June 30. This means that children born in June have *cutoff* = 0, those born in July *cutoff* = 1, and so on.
Source: SOEP v37, own calculations. N = 20,893

Figure 2.2. Discontinuity in age at school starting

As shown in Figure 2.2, in our sample of interest in the SOEP, where we can see the exact schooling histories of individuals, the probability of going to school at age six drops substantially after the cutoff. Since the Microcensus is cross-sectional and does not contain complete educational histories, to test our assumption we circumvent this limitation by focusing on respondents for whom we can observe their contemporaneous school placement and grade level, allowing us to reconstruct their educational trajectory, as we did for Figure 2.A.2. Similarly for the Microcensus, Figure 2.A.2 shows that the shift in probability is substantial.

Then, we go a step beyond and also investigate other pivotal moments in the individuals' schooling period, to see whether correction of the cutoff takes place after starting school. For instance, if students switching from elementary to secondary school around the cutoff are equally likely to be either still in grade four or already in grade five, this would mean that they correct their school trajectory after having entered school, either by repeating or skipping a year. If this adjustment would be correlated with our outcome of interest, this could bias our results.

Our preferred fuzzy RDD specification is as follows:

$$y_{isc} = \beta_1 Before_i + \beta_2 Male_i + \beta_3 Before_i \times Male_i + \beta_4 MonthBirth_i \\ + \gamma_c^{cohort} + X_i + \epsilon_{isc} \quad (2.1)$$

where y_{isc} is the outcome of interest for individual i, in state s and cohort c; $Before_i$ takes value 1 for students born before the cutoff, 0 otherwise; $Male_i$ is a dummy variable that takes value 1 if i is male, 0 otherwise. $MonthBirth_i$ is the running variable, X_i a vector of individual controls, and we further include year of birth fixed effects. All standard errors are clustered at the state level. The coefficients β_1 and β_3 are our main coefficients of interest, capturing the effect for women (β_1) and for men ($\beta_1 + \beta_3$).

A few features of this specification differ from "standard" RD design. Rather than a local linear regression with data-driven bandwidth selection – the approach most common in RDD settings with a continuous running variable – we work with a fixed window of two months on either side of the cutoff and control for month of birth linearly. This choice is motivated by the discrete nature of our running variable: month of birth takes at most twelve distinct values, and within our window of interest only four. With so few mass points, conventional bandwidth selection procedures have little traction, and estimating flexibly-varying trends on each side of the cutoff is not feasible. The two-month window reflects a deliberate tradeoff: children born within two months of the cutoff are highly comparable in age and developmental stage, whereas expanding the window progressively reduces this comparability. We therefore prefer to anchor identification in this near-birthday comparison rather than in a data-driven bandwidth, and show in the Appendix (Table 2.A.5) that our results are robust to alternative window widths, running variable specifications, and a local linear regression approach.

Since not all children comply with the cutoff rule, equation 2.1 estimates an intention-to-treat (ITT) effect rather than a local average treatment effect (LATE). The ITT captures the reduced-form impact of being born just before the cutoff – and thus being assigned to earlier school entry – on occupational segregation, without scaling by the share of compliers. We prefer this interpretation for two reasons. First, as discussed above, the presence of two-sided non-compliance makes the monotonicity assumption required for a LATE interpretation difficult to defend formally, even if we argue its practical relevance is limited. Second, from a policy perspective, the ITT directly answers the question of whether shifting the school entry cutoff – as the reforms we exploit did – produces changes in gender occupational segregation, which is the relevant counterfactual for policy evaluation.

Table 2.2. First stage RDD

	SOEP	Microcensus				
	1996-2004	2007-2011	1998-2002		1997-2001	
	Starting school at 6		Grade 9	VET	Grade 10	Grade 11
	(1)	(2)	(3)	(4)	(5)	(6)
Before	0.44***	0.40***	−0.019	0.0011	−0.18***	0.048
	(0.072)	(0.084)	(0.024)	(0.0058)	(0.031)	(0.040)
Male	0.027	−0.015	0.014	0.0030	0.010	−0.060**
	(0.052)	(0.017)	(0.024)	(0.0023)	(0.012)	(0.023)
Before × Male	−0.037	−0.034*	−0.015	0.0051	−0.0048	−0.049*
	(0.059)	(0.017)	(0.023)	(0.0047)	(0.018)	(0.027)
Month of birth	−0.044***	−0.042***	0.023***	−0.0052***	0.027***	−0.013**
	(0.009)	(0.0098)	(0.0036)	(0.0015)	(0.0042)	(0.0050)
Year of birth FE	✓	✓	✓	✓	✓	✓
Baseline mean	0.21	027	0.81	0.02	0.37	0.41
N	3,088	5,535	2,774	7,123	4,310	4,310
Adj. R^2	0.478	0.373	0.046	0.016	0.161	0.020
F Statistic	287.4	354.3	47.2	6.86	150.6	16.3

Notes: The table shows the following results: Column (1) presents regression results of the probability change of going to school at age six around the school entry cutoff date using SOEP data. The baseline mean is the probability of going to school at age six when being born after the cutoff. Columns (2)-(6) use Microcensus data. Given the cross-sectional nature of the Microcensus data, we observe only individuals' current educational status and outcomes; thus, for this analysis we restrict our sample to respondents for whom we can identify both their current school type and grade level at the time of the survey. In column (2) is estimated the effect of being born before the cutoff on the probability of starting school at age six. In column (3) the dependent variable is the probability of being still in grade nine for the subsample of students attending low track secondary school. In column (4) we estimate the effect of being born before the cutoff on the probability of being already in VET, conditional on having a low track school leaving degree; and in columns (5)-(6) the dependent variable is the probability of being in grade ten and eleven for the subsample of students enrolled in a high track secondary school. The negative coefficient of the variable *Before* in columns (3)-(5) indicates that students born *before* the cutoff are *less* likely to be still in grade nine/ten at the age of sixteen/seventeen. The opposite is true for columns (4) and (6), albeit noisier estimates. To ease of interpretation, the baseline gives the mean for women born after the cutoff (for which *Before* = 0). Standard errors are bootstrapped and clustered at the state level. SOEP v37 and Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We analyze three pivotal moments of the students' life: (i) school entry (six years old); (ii) for low track secondary school students, attending grade nine vs. starting VET (sixteen/seventeen years old); and (iii) for high track secondary school students, attending grade ten vs. grade eleven (seventeen/eighteen years old). For each regression, we select the Microcensus sample as to consider only those individuals who, for each wave, would have had the right age. For example, to estimate the effect of being born before the cutoff on the probability of starting school at age six, we restrict our sample in survey wave 2014 to individuals born between January and December 2007. Since the survey took place in early 2014, we would expect children born *before* July 2007 to be enrolled in school, whereas those born in July or after would not be enrolled. While this sample does not directly coincide with our main analysis sample, they are of very similar age and do not show substantially different trends in schooling attendance, suggesting that estimations for our cohorts of interest should not differ by much.

Table 2.2 presents these estimates. Columns (1)-(2) show the regression on school entry compliance at six years old for the SOEP (our cohorts of interest) and the Microcensus (younger cohorts not yet in the labor market), respectively; columns (3)-(4) show respectively the probability of being still in grade nine and being in VET at sixteen and seventeen years old, conditional on being enrolled in/having a school leaving degree from a low track secondary school, respectively; and columns (5)-(6) show the probability of attending grade ten and grade eleven, conditional on being enrolled in a high track secondary school, at seventeen and eighteen years old respectively.

Being born before the cutoff significantly increases the probability of going to school when aged six in the SOEP sample. As the interaction coefficient suggests, there are no differences by gender, so parents are not more inclined to anticipate or postpone children's school entry according to their gender. The F-statistic of 287 implies a high relevance of the instrument in determining the assignment rule. For the Microcensus sample, children born before the cutoff are 40 p.p. more likely to be enrolled in primary school when aged six, a near-identical estimate to the one from the SOEP. The coefficient of the interaction term *Before* × *Male* is negative and weakly significant, meaning that boys could be slightly more redshirted by parents than girls, thus starting school later even if being born before the cutoff. The results are robust when we perform a local linear regression (see Table 2.A.1).

Moving on to the next steps in schooling histories, the effect is quantitatively similar for students attending a low track secondary school, albeit not statistically significant when controlling for month of birth as running variable. We show in Table 2.A.4 that controlling for different combinations of year-month of birth yields robust and qualitatively similar coefficients to the ones shown in Table 2.2, indicating that individuals born before the cutoff are still relatively more likely to finish school earlier.

Conditioning on attending a high track secondary school, they are also 18 p.p. less likely to be still in grade ten at the age of seventeen. Boys are also less likely to be in grade eleven at the age of eighteen, confirming the persistence of the effect we observe at school start, so that those who enter school *later* appear to remain in school one year more.

Table 2.3. Age at school exit by school leaving degree in the SOEP

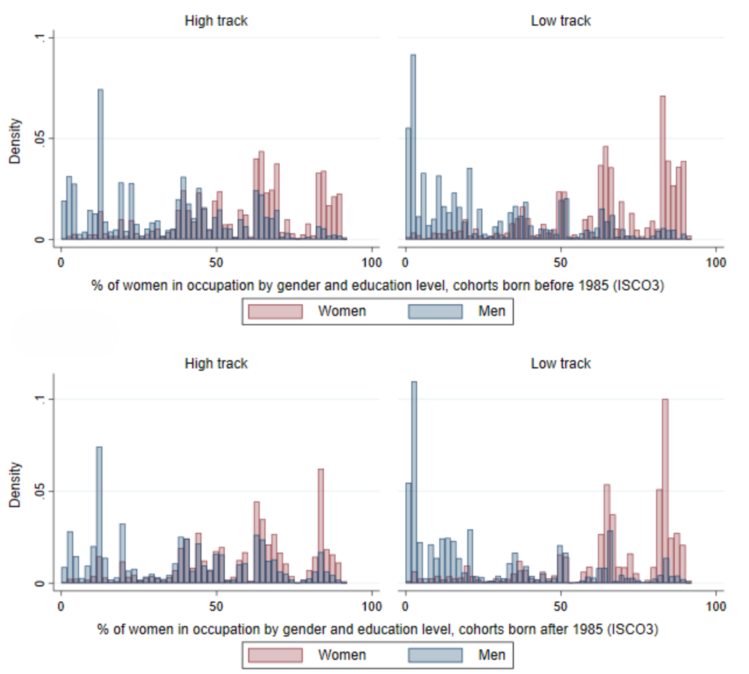
	Age at School Exit		
	Low track	Vocational Track	High Track
	(1)	(2)	(3)
Before	−0.687***	−0.616**	−0.582**
	(0.233)	(0.252)	(0.288)
Male	0.086	0.164*	−0.232**
	(0.170)	(0.098)	(0.100)
Month of birth	−0.014	0.030	0.025
	(0.024)	(0.045)	(0.033)
Before × Male	0.181	0.058	0.073
	(0.265)	(0.133)	(0.166)
Year of birth FE	✓	✓	✓
State FE	✓	✓	✓
Baseline mean	16.65	17.49	19.37
N	2,046	3,858	7,779
Adjusted R^2	0.074	0.109	0.095

Notes: The table shows how age at school exit changes when being born before school entry cutoff dates. The baseline mean is the average age at school exit of individuals born after the cutoff (dependent variable). The regression is estimated separately by school leaving degree and with State FE to account for different secondary school tracks' duration. If there were perfect compliance to the rules, the difference would be exactly one year. The coefficient of *Before* shows that, despite non-compliers, those repeating or skipping one year, the difference is still close to one. To ease of interpretation, the baseline gives the mean for women born after the cutoff (for which *Before* = 0). Robust standard errors are clustered at the state level in parentheses. SOEP v37, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, our assumption implies that age at school entry affects age at school exit, i.e. age at labor market entry. We test this by regressing age at school exit on whether individuals are born before or after the cutoff in the SOEP, since we do not have comparable reliable information in the Microcensus. Table 2.3 shows separate estimates by school track choice. If there were 100% compliance and no adjustment (skipping or repeating a year) after school start, the coefficient should be equal to minus one. Despite lower actual compliance and the possible adjustments later, our estimates suggest that individuals born before the cutoff are still likely to finish school 2/3 of a year sooner than those born after, meaning that school eligibility has a substantial impact on age at labor market entry.

Our dependent variable of interest y_{isc} is a dummy variable equal to 1 when individual i, in state s and cohort c works in a gender segregated occupation, 0 otherwise. We construct our measure of segregation by looking at the distribution of all the occupations in the whole of the Microcensus sample (see Figure 2.3). We define an occupation as male (female)-segregated if 10% or less of the workers are women

(men). We show that our results are robust to choosing different segregation thresholds (Figure 2.A.6).



Notes: The figure shows the distribution of workers in occupations in the Microcensus by gender, cohort and educational level. Occupations are arranged along the x-axis according to their female composition, from 0% (exclusively male) to 100% (exclusively female). In red is shown the distribution of women; in blue the distribution of men. On top, the results for cohorts born before 1985 (N = 1,462,307) are presented and in the bottom panel are the cohorts born after 1985 (N = 414,078). On the left are the results for people who have a high track school leaving degree and on the right for people who have a low track school leaving degree.
Source: Microcensus 2014-19, own calculations.

Figure 2.3. Distribution of workers in occupations (ISCO3) by gender, educational level, and birth cohort

2.3.2 Descriptive statistics

Table 2.4 presents summary statistics of the entire Microcensus sample, from which we derive our measure of segregation, in columns (4)-(6); and of our relevant cohorts affected by the reforms (1996-2004) in columns (1)-(3). Since we are especially interested in individuals graduating from low track secondary school, results are provided separately for this subsample as well in columns (1) and (4).

Table 2.4. Summary statistics

	Relevant cohorts			All sample		
	Low track mean/sd (1)	High track mean/sd (2)	Total mean/sd (3)	Low track mean/sd (4)	High track mean/sd (5)	Total mean/sd (6)
Gender	0.436	0.536	0.487	0.513	0.501	0.508
	(0.496)	(0.499)	(0.500)	(0.500)	(0.500)	(0.500)
Age	19.200	20.560	18.370	46.330	43.580	45.060
	(2.002)	(1.576)	(2.396)	(13.640)	(12.950)	(13.480)
Had already worked	0.222	0.221	0.095	0.853	0.727	0.773
	(0.416)	(0.415)	(0.293)	(0.354)	(0.446)	(0.419)
Works full time	0.907	0.623	0.747	0.717	0.748	0.730
	(0.290)	(0.485)	(0.435)	(0.450)	(0.434)	(0.444)
German citizenship	0.924	0.966	0.944	0.920	0.940	0.916
	(0.349)	(0.312)	(0.343)	(0.327)	(0.321)	(0.341)
Completed VET	0.250	0.101	0.104	0.798	0.412	0.603
	(0.433)	(0.301)	(0.306)	(0.402)	(0.492)	(0.489)
University degree	0.000	0.046	0.011	0.001	0.470	0.206
	(0.007)	(0.209)	(0.105)	(0.030)	(0.499)	(0.405)
Nonmissing occupation	0.663	0.377	0.314	0.718	0.787	0.736
	(0.473)	(0.485)	(0.464)	(0.450)	(0.409)	(0.441)
% women in occupation	0.417	0.526	0.454	0.439	0.472	0.454
	(0.326)	(0.273)	(0.313)	(0.305)	(0.256)	(0.284)
25% segr. occupation	0.666	0.486	0.605	0.589	0.418	0.508
	(0.472)	(0.500)	(0.489)	(0.492)	(0.493)	(0.500)
10% segr. occupation	0.287	0.093	0.222	0.233	0.101	0.172
	(0.452)	(0.290)	(0.416)	(0.423)	(0.301)	(0.377)
N	89,758	70,416	288,537	1,271,320	1,059,924	2,418,624

Notes: This table presents summary statistics. Columns (1)-(3) report summary statistics for the cohorts of interest (1996-2004). Results are presented for low and high track school graduates separately, as well as pooled together. Columns (4)-(6) report summary statistics for all sample of employed individuals from which we derive our measure of occupational segregation. Microcensus 2014-19, own calculations.

While our relevant cohorts are significantly younger (as evidenced by the age and related educational/work experience differences), most other characteristics are notably similar. The comparable values for gender distribution, citizenship, occupational segregation measures, and percentage of women in occupations suggest that our cohorts of interest are similar to the whole sample, with residual differences attributable to the age gap.

Table 2.5 shows how workers in segregated occupations are distributed by gender and school leaving degrees. Segregation is much more prominent for individuals graduating from low track secondary school: one in two men graduating from the low track works in occupations with 10% of women or less, whereas only one in five high track graduates are expected to do so. Furthermore, we see that segregation on the labor market affects men and women differently: across school leaving degrees, only one in seventeen women works in a female-segregated occupation, whereas more than one man in four is employed in a male-segregated occupation.

Table 2.5. Occupational segregation by gender and high school leaving degree

	Share of workers in male occupations			Share of workers in female occupations		
	Male	Female	Total	Male	Female	Total
High track	0.22	0.03	0.12	0.02	0.08	0.05
Low track	0.50	0.04	0.32	0.01	0.15	0.06
Total	0.43	0.03	0.26	0.01	0.13	0.06

Notes: The table presents the share of workers in male-dominated occupations on the left side and the share of workers in female-dominated occupations on the right side, divided by educational track. An occupation is defined as male (female)-dominated if 10% or less of the workers are women (men). Microcensus 2014-19, own calculations. N = 2,331,244.

Figure 2.3 illustrates the distribution of workers in occupation by gender, school track and age. Gender segregation is fairly persistent over time: if we plot the gender composition of occupations, we see that little difference occurs if we consider older cohorts born before 1985, or younger ones born after 1985; the main difference is still given by the school leaving degrees. Individuals graduating from the high track select into occupations more at the center of the distribution (where female share is more or less close to 50%), whereas those from the low track are much more concentrated around the extremes, i. e. men (women) select into occupations with a very low share of women (men).

Finally, Table 2.A.3 in the Appendix shows summary statistics for our cohorts of interest by treatment status. We see that the two groups do not differ substantially in their main characteristics.

2.3.3 Effect of age at school entry on gender segregation in the labor market

Table 2.6. RDD model: The effect of age at school entry on labor market segregation

	Segregated occupation (25%)		Segregated occupation (10%)	
	All sample	Low track sample	All sample	Low track sample
	(1)	(2)	(3)	(4)
Before	-0.020^{*}	-0.016	-0.0067	-0.00080
	(0.011)	(0.013)	(0.014)	(0.020)
Male	0.18^{***}	0.19^{***}	0.30^{***}	0.36^{***}
	(0.012)	(0.014)	(0.019)	(0.014)
Before × Male	0.044^{**}	0.056^{**}	0.025^{*}	0.030^{**}
	(0.015)	(0.021)	(0.013)	(0.014)
Month of birth	-0.0034	-0.000098	-0.00067	0.0013
	(0.0073)	(0.0068)	(0.0068)	(0.010)
Year of birth FE	✓	✓	✓	✓
Baseline mean	0.52	0.55	0.07	0.08
Controls	✓	✓	✓	✓
N	17,294	11,958	17,294	11,958
Adjusted R^2	0.045	0.054	0.136	0.157

Notes: This table presents our main RDD estimates. In columns (1)-(2) is shown the effect of being born before the cutoff on the probability of working in an occupation with 25% or less other-gendered workers, and 10% or less in columns (3)-(4). We control for month of birth as running variable and year of birth FE. Further controls include migration background, ZIP code area (NUTS 3), interview quarter year, G8 reform. To ease of interpretation, the baseline gives the mean for women born after the cutoff (for which *Before* = 0). Robust standard errors are clustered at the state level in parentheses. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6 displays the 'reduced form' estimates of our fuzzy RDD using Microcensus data. Our preferred specification includes two months before and two after the cutoff. While the regressions presented here include month of birth as continuous variable, in Table 2.A.5 we show that the results remain identical when including month of birth FEs, as well as controlling for different combinations of year-month of birth FEs.

The dependent variable in column (1) and (2) is a dummy variable equal to 1 if the individual works in an own-gender segregated occupation (occupation that has 75% of own-gender workers, or more); for columns (3) and (4), we use a 10% indicator (occupation has 90% of own-gender workers, or more). In Figure 2.A.6, we show that our results are robust to choosing different segregation thresholds, as well as using a continuous measure of segregation on the left hand side of the equation. Column (1) and (3) show the results across all individuals affected by the reforms, whereas column (2) and (4) present the estimates for the subsample of individuals holding a low track school leaving degree only. The coefficient associated with the variable *Before* identifies the treatment effect for women, whereas the effect for men is given by the sum of the coefficients from the variable *Before* and the interaction term *Before* × *Male*, while the coefficient for *Male* identifies baseline differences in occupational segregation by gender.

Estimates for men yield positive and significant coefficients in both groups. On average, men graduating from the low track born before the cutoff (who then enroll to school when aged six, thus *younger* than their counterparts) are on average 3 percentage points more likely to select into a heavily male-segregated occupation, an increase of 7%.

For women instead, the treatment effect is not significant and very close to zero. This means that shifting the age at labor market entrance for women within this age interval does not seem to produce a substantial effect on their choice of occupation in terms of gender segregation.

2.4 Robustness checks

In this section we discuss potential concerns, mechanisms, robustness checks and explanations for our results.

2.4.1 Effect on school track choice

One concern is that individuals who go to school one year earlier are not only younger at the school start or exit, but also at each pivotal moment of their schooling period. Thus, every decision they make regarding school, both at the extensive and the intensive margin, might be affected by their younger age. If, for example, younger, less mature children are also more likely to select into a low track secondary school, and this in turn affects their occupational choices in terms of gender segregation, then a valid concern would be that the effect that we capture does not only reflect age, but it might be mediated by the earlier track choice.

While studies looking at international comparisons find in general a small, but persistent negative effect for early tracking on younger individuals, evidence from Germany shows that, on average, tracking decisions do not affect any long term labor market outcome (Dustmann, Puhani, and Schönberg, 2017). This null effect seems to be driven by the flexibility of the system (Mühlenweg and Puhani, 2010) in which at least one student out of four at the end of the schooling period has changed track (Bildungsberichterstattung, 2020). Therefore, it is unlikely that we are capturing the only labor market outcome influenced by earlier tracking, assuming that students who were assigned to tracks earlier *because* of the reform do not subsequently adjust their educational trajectory, contrary to their peers.

To check for this possibility, we would ideally like to investigate whether tracking is affected by earlier school entry in our sample of interest. However, due to the cross-sectional nature of the Microcensus data, we can only look at school leaving degrees once finished school and we cannot account for any track change within the schooling period. To test this, we employ two strategies: (i) we follow the analysis of Dustmann, Puhani, and Schönberg (2017) with the cohorts most close to ours, but not directly affected by the reforms, to see if also younger generations' age-at-tracking does not affect school leaving degrees; (ii) we repeat this by limiting our sample to the 2019 survey wave, so that the first treated cohort with a sizable sample (1999) is old enough to reasonably have already obtained a school leaving degree.

Table 2.7. RDD model for school leaving degrees

	cohorts 1989-1994			cohort 1999		
	Low track	Technical high track	High track	Low track	Technical high track	High track
	(1)	(2)	(3)	(4)	(5)	(6)
Before	−0.0042	0.013	−0.017	−0.0083	0.0036	−0.063
	(0.012)	(0.0098)	(0.010)	(0.046)	(0.019)	(0.037)
Male	−0.0019	0.0057	-0.067^{***}	0.062^{***}	0.0044	-0.098^{***}
	(0.014)	(0.0061)	(0.0059)	(0.017)	(0.018)	(0.017)
Before × Male	0.0076	−0.0093	0.00070	−0.0034	−0.014	−0.019
	(0.0070)	(0.0055)	(0.0058)	(0.023)	(0.021)	(0.032)
Month of birth	0.00015	−0.0013	0.0025	−0.00015	−0.0026	-0.042^{**}
	(0.0049)	(0.0029)	(0.0031)	(0.017)	(0.0080)	(0.019)
Year of birth FE	✓	✓	✓	–	–	–
Baseline mean	0.31	0.13	0.44	0.32	0.09	0.47
Controls	✓	✓	✓	–	–	–
N	87,599	87,599	87,766	2,219	2,219	2,225
Adj. R^2	0.055	0.020	0.109	0.002	-0.002	0.012

Notes: This table shows the RDD of probability of getting low track and (technical) high track school leaving degrees in columns (1)-(3) for early vs. late school goers in the before-treatment period, cohorts 1989-1994 (two months around the cutoff); and for individuals born before vs. after the cutoff for the cohort 1999 in columns (4)-(6). Interval chosen as to have cohorts near in time to our sample of interest, and so that, in the respective survey wave, the youngest ones were aged 20 and thus had very likely finished school. To ease of interpretation, the baseline gives the mean for women born after the cutoff (for which *Before* = 0). Robust standard errors are clustered at the state level in parentheses. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As found by Dustmann, Puhani, and Schönberg (2017), also in more recent cohorts earlier school entry does not coincide with a different probability of getting different school leaving degrees. Individuals born before the cutoff, who then enroll in school *earlier*, do not display any differing school trajectory from their later sorted counterparts. The effect is absent for both genders, although their baseline differences in probability of getting specific school degrees: men are on average 7-10 p.p. less likely to obtain a high track school leaving degree (16-19%) and 6 p.p. more likely to get a low track school leaving degree with respect to women (although only in the case of the 1999 cohort). This is in line with the evidence that children who start school earlier and who then might get sorted into a lower track than their 'optimal' one, eventually do not show any substantial difference in schooling outcomes. That is, thanks to the flexibility of the German schooling system, they still have the same chance of getting a (technical) high track degree.

Still, data limitations do not allow us to exclude that within-track sorting affects our outcome of interest, or that, while exiting school with the same degrees, 'wrong' sorting could be one of the drivers.

2.4.2 Birth timing and month of birth effects

Another potential concern with our identification strategy is the possibility of systematic differences in birth patterns around the school entry cutoff dates. If parents strategically timed births to take advantage of the cutoff rules, this could lead to non-random sorting of individuals around the cutoff.

However, as shown in Figure A.4, the data does not indicate any unusual patterns or discontinuities in birth rates around the school entry cutoff dates. While there is evidence of some seasonal variation in birth rates, this seasonality appears to be orthogonal to the school entry cutoff dates, as indicated by the smooth distribution of births across months without any sharp changes at the cutoff points. Furthermore, the Frandsen test results presented in Table A.2 provide additional reassurance. This test is designed to detect any unusual heaping or nonlinearity in the distribution of a discrete running variable (birth month) around the cutoff. For the relevant cohorts (1996 to 2004), the test does not find significant evidence of manipulation, with p-values above conventional significance levels even at lower levels of the K parameter that are more sensitive to small deviations from linearity, especially for the cohorts affected by the reforms.

Nonetheless, it could be that children born in particular months have characteristics correlated with our outcome of interest. To control for this, we augment our analysis with a difference-in-difference design (DiD), in which we apply a triple-difference to capture (i) the difference between states that are treated and states that are not; (ii) the difference between individuals born in different months, holding year of birth constant; and (iii) the difference between those born in the same months, but in different years, holding state constant.

We adapt our RDD specification by controlling for state, month of birth and cohort fixed effects:

$$y_{igc} = \beta_1 D_i + \beta_2 Male_i + \beta_3 Male_i \times D_i + \gamma_c^{cohort} + \delta_g^{state \times month} + X_i + \epsilon_{igc} \qquad (2.2)$$

where y_{igc} is an indicator for gender segregation of the occupation of individual i, born in state-month of birth combination g and in cohort year c; D_i indicates whether individual i is born in the states, years and months affected by the school entry reforms, i. e. is equal to 1 for children going to school one month *earlier* to their respective counterparts. Robust standard errors are clustered at the state × month level in our preferred specification, and we control for a vector of individual characteristics X_i. The coefficients β_1 and β_3 capture the impact of age at school entry on occupational segregation for women and men, conditional on fixed differences across state-months of birth and fixed differences across cohorts. Thus, the TWFE estimator allows us to estimate the ITT effect. Fixed differences across states or seasonality effects of month of birth cannot drive our results, because they are captured by $\delta_g^{state \times month}$ and we exploit the variation over time within each state × month of birth combination. Likewise, cohorts effects across states and months of births are captured by γ_c^{cohort} fixed effects. Our results are robust to changing the specification with state × cohort and month of birth fixed effects.

Our causality claim holds under the parallel trends assumption only. This means that, in the absence of treatment, it should be plausible to assume that the two groups' main outcomes would have had similar trajectories. While we never observe counterfactuals, we provide suggestive evidence of similarity between labor market outcomes and educational paths between treated and control states. Figure 2.A.7 shows, for example, that wages, school leaving degrees and share of workers employed in segregated occupations do not display different trends for the cohorts born in the period just before the reforms apply. Also, to the best of our knowledge, no competing intervention was adopted affecting the age group and states of our sample of interest.

Furthermore, we go a step beyond and compare not simply individuals born in treated and control states, but we define our level of comparison at the state × month of birth. This implies that a violation of the parallel trends clause, in order to affect our results, should be year-specific across all states (or at least within treated and control states), but differentially affect individuals born in the months of the school entry reforms. Similarly, in order for us to pick up a spurious effect given by seasonality of birth, we should assume that this month-specific trend differs precisely across treated and control states.

We present results for the DiD specification in Table 2.8. As in the previous model, we estimate the regression on all sample and on the subsample of low track graduates only, in column (1) and (2) respectively. In column (3) and (4) we repeat the same regression, but we control for state × year and month of birth fixed effects; the results remain unvaried.

Table 2.8. Difference-in-Difference estimators: the effect of age at school entry on labor market segregation

	Segregated occupation (10%)			
	All sample	Low track	All sample	Low track
	(1)	(2)	(3)	(4)
D	−0.022	−0.0031	−0.022	−0.011
	(0.016)	(0.019)	(0.020)	(0.023)
Male	0.28***	0.32***	0.28***	0.32***
	(0.0047)	(0.0061)	(0.0047)	(0.0061)
D × Male	0.086***	0.056***	0.089***	0.062***
	(0.019)	(0.021)	(0.019)	(0.022)
Year of birth FE	✓	✓	–	–
State × month of birth FE	✓	✓	–	–
Month of birth FE	–	–	✓	✓
State × year of birth FE	–	–	✓	✓
Controls	✓	✓	✓	✓
N	29,598	21,193	29,598	21,193
Adj. R^2	0.134	0.143	0.127	0.139

Notes: The table shows the effect of the treatment (being born *before* the cutoff in states where school entry rules changed) on the probability of working in a gender-segregated occupation, defined as a dummy variable equal to 1 if less than 10% of workers have a different gender. The coefficient for the treatment variable *D* gives the effect for girls of going to school one year *earlier* than girls born in the same month in previous years in the same state, and in the same month and year but in a different state. The coefficient of the variable *Male* is the baseline difference in occupational segregation by gender. The interaction term *D* × *Male* gives the treatment effect for boys. Robust standard errors are clustered at the state × month of birth level in parentheses for columns (1) and (2), and at the state × year of birth level in columns (3) and (4). Controls include migration background, ZIP code area (NUTS 3), interview quarter year. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The variable D refers to the treatment status, so $D = 1$ means that individuals went to school one year *earlier* than their counterparts in previous years or in the control group. The estimates are in line with the RDD model and show that younger men select themselves into more male-segregated professions. The effect is qualitatively similar to the RDD estimation: low track graduates are on average 5.6 p.p. more likely to choose a segregated profession; across all sample, the effect is equal to 8.6 percentage points.

For women, the effect is not significant and, if ever, slightly negative.

The results indicate that boys who enter and exit school one year earlier are more likely to choose gender-typical occupations, while girls' decisions remain unaffected by the younger age. We consider several mechanisms that could explain these results.

2.4.3 Mirroring parental labor market choices

One possible explanation of this effect could be that individuals who start school earlier, while having the same amount of education as their peers, have more limited maturity and are less exposed to experiences different from the ones revolving around the family environment. This mirroring effect could lead them to emulate the gendered patterns prevalent in their parents' professions, thus perpetuating occupational segregation.

2.4.4 Stereotypes and norms

Since we observe a different effect of age at school entrance by gender, it is also possible that this age represents different maturity levels and different crucial periods for girls and boys. One possibility is that stereotypes onset on girls is earlier, so that it does not matter whether decision is anticipated at fifteen or sixteen because it was already made earlier. Hence, the crucial period for making this type of decision might differ by gender. A potential next step is to collect survey data on when adolescents make occupational decisions and whether or not this age differs by gender.

2.4.5 Relative age effect

As mentioned in section 2.4.1, the reform shifts the age of treated individuals throughout the entire schooling period: this means that they are also on the left tail of the age distribution among their classmates. The literature shows that older individuals in the classroom are more self-confident (Page, Sarkar, and Silva-Goncalves, 2019), and the effect appears to be lasting in life (Du, Gao, and Levi, 2012). In most states, the reform shifts school entry for one month of birth, so that we actually compare individuals who are at the extreme tails of the distribution, and this could affect our results if more self-confident individuals were also more likely to select into more gender-atypical occupations. However, in some states, the reform shifted the school cutoff entry by many months directly: for example, in Berlin, starting already from the first treated cohort, the cutoff was shifted from June 30th to December 31st abruptly. This means that we can compare there individuals who are in the middle, and not at the tails of the age distribution in class.

Table 2.9 shows the results of the RDD specification for the Berlin subsample. We see that the effect is qualitatively comparable, although somewhat noisier given the small sample size. This would suggest that the results are not driven by the relative age in class of the individuals schooled earlier.

Table 2.9. Relative age effect: RDD in state of Berlin

	Segregated occupation (25%)		Segregated occupation (10%)	
	All sample	Low track sample	All sample	Low track sample
	(1)	(2)	(3)	(4)
Before	−0.0043	−0.097	−0.10*	−0.25**
	(0.099)	(0.14)	(0.061)	(0.10)
Male	0.068	−0.067	0.16***	0.14*
	(0.075)	(0.10)	(0.047)	(0.074)
Before × Male	−0.038	0.032	0.10*	0.19**
	(0.092)	(0.12)	(0.061)	(0.091)
Month of birth	−0.0051	−0.0085	−0.024*	−0.051**
	(0.019)	(0.025)	(0.013)	(0.020)
Year of birth FE	✓	✓	✓	✓
Baseline mean	0.58	0.77	0.11	0.23
N	524	285	524	285
Adj. R^2	-0.002	-0.019	0.113	0.146

Notes: This table presents the RDD specification for the Berlin subsample, where the cutoff was shifted by six months directly starting from the cohort 1999. To increase the sample size to get meaningful estimates, we consider four months around the cutoff. The coefficient of *Before* gives the effect of going to school before the school entry cutoff date for girls. The coefficient of *Male* is the baseline difference in occupational segregation by gender. The interaction term *Before* × *Male* gives the effect of going to school earlier for boys. To ease of interpretation, the baseline gives the mean for women born after the cutoff (for which *Before* = 0). Robust standard errors are clustered at the state level in parentheses. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5 Conclusion

This paper examines persistent gender-specific occupational choices, focusing on age at labor market entry as a potential cause of occupational gender segregation. Using Germany's early-tracking educational system, we analyze a series of school entry cutoff reforms that create quasi-exogenous variation in labor market entry age. This natural experiment allows us to identify the causal relationship between age at entry and gender-specific occupational outcomes, particularly within vocational education pathways.

Our findings show that entering the labor market at a younger age significantly increases the likelihood for men to choose male-dominated occupations, whereas women's occupational sorting appears unaffected by entry age. This asymmetric response suggests that younger men may default to stereotypically male occupations when required to make early career decisions. These default choices are likely influenced by the salience of gender identity during adolescence — a period when gender identity tends to solidify (Ramaci et al., 2017). Early decision-making may thus contribute to the persistence of gender norms by locking individuals into same-gendered peer groups that perpetuate traditional roles. In magnitude terms, the estimated effect for men implies roughly a 7% increase in entry into male-segregated occupations, which is economically meaningful given the size of apprenticeship cohorts and the persistence of initial occupational matches.

Occupational choice represents one of the most significant decisions individuals face, with people spending over 80,000 hours working across a typical career. Initial occupational decisions have been shown to exert durable effects on health behaviors and well-being (Kelly et al., 2014). Moreover, switching careers becomes increasingly difficult as individuals invest more in their chosen professions, making regret over initial choices a persistent experience (Budjanovcanin, Arinto Martins Rodrigues, and Guest, 2020). Research indicates that men are more likely than women to report work-related regret (Morrison and Roese, 2011), underscoring the potential for early, default-driven decisions to yield enduring dissatisfaction.

From a policy perspective, our results highlight the equity risks of early tracking systems, especially those that require students to make career-defining decisions at a young age. Since age at labor market entry can be influenced by relatively low-cost interventions — such as adjusting school entry cutoffs — our findings point to feasible policy levers. However, such reforms must be approached with caution, as later school entry may have unintended consequences for educational and developmental outcomes. The labor market implications are potentially broader than occupational composition alone: if earlier entry shifts more young men into narrowly male, blue-collar training occupations, it may also increase exposure to segments with higher cyclical unemployment and displacement risk. In that sense, even moderate effects on segregation can translate into non-trivial differences in subsequent unemployment risk and employment stability.

Future research could expand upon our findings by investigating whether delayed occupational choices lead to reduced gender wage gaps, increased career mobility, or enhanced well-being. Comparative studies could also assess the validity of our results in other educational systems, both tracked and untracked. Additionally, a closer examination of how social context, peer effects, and parental expectations influence the relationship between age and occupational choice could enhance our understanding of the developmental origins of gendered labor market trajectories.

Although the current study cannot assess long-term labor market outcomes as the reforms affect individuals still in the early stages of their careers, future research could also examine how occupational segregation interacts with structural labor market changes. Many male-dominated vocational occupations are susceptible to automation, which could put men at a disadvantage if they enter such roles early and remain in them. This underscores the potential long-term risks of early occupational sorting and emphasizes the importance of timing in career decision-making.

References

Allen, Jeremiah, and Roger Barnsley. 1993. "Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection." *Journal of Human Resources* 28 (3): 649–59. [74]

Antecol, Heather, and Deborah A. Cobb-Clark. 2013. "Do psychosocial traits help explain gender segregation in young people's occupations?" *Labour Economics* 21: 59–73. [75]

Attanasio, Orazio P, and Katja M Kaufmann. 2017. "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations." *Journal of Economic Behavior & Organization* 140: 35–55. [75]

Barua, Rashmi, and Kevin Lang. 2016. "School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect." *Journal of Human Capital* 10 (3): 347–76. [82]

Berthold, Barbara. 2008. "Einschulungsregelungen und flexible Eingangsstufe." *Recherche für den Nationalen Bildungsbericht,* 30. [79, 115]

Bildungsberichterstattung. 2020. *Bildung in Deutschland 2020: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt.* wbv Media GmbH & Co. KG. [77, 92]

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2011. "Too young to leave the nest? The effects of school starting age." *Review of Economics and Statistics* 93 (2): 455–67. [82]

Budjanovcanin, Amina, Rodrigo F. Arinto Martins Rodrigues, and Daniel Guest. 2020. "A career with a heart: Exploring occupational regret." *Journal of Managerial Psychology* 34 (3): 156–69. [99]

Chen, Kelly, Nicole Fortin, and Shelley Phipps. 2015. "Young in class: Implications for inattentive/hyperactive behaviour of canadian boys and girls." *Canadian Journal of Economics/Revue Canadienne D'ÉConomique* 48 (5): 1601–34. [74]

Cortes, Patricia, and Jessica Pan. 2018. "Occupation and gender." *Oxford Handbook of Women and the Economy,* 425–52. [75]

Delfino, Alexia. 2024. "Breaking gender barriers: Experimental evidence on men in pink-collar jobs." *American Economic Review* 114 (6): 1816–53. [75]

DIW. 2022. *Socio-economic panel (SOEP), data for years 1984-2020, SOEP-Core v37, EU Edition.* [80, 81]

Du, Qianqian, Huasheng Gao, and Maurice D Levi. 2012. "The relative-age effect and career success: Evidence from corporate CEOs." *Economics Letters* 117 (3): 660–62. [74, 97]

Dustmann, Christian, Patrick A. Puhani, and Uta Schönberg. 2017. "The long-term effects of early track choice." *Economic Journal* 127 (603): 1348–80. [92, 93]

Faust, Gabriele. 2006. "Zum Stand der Einschulung und der neuen Schuleingangsstufe in Deutschland." *Zeitschrift Für Erziehungswissenschaft* 9 (3): 328–47. [78, 79, 112, 115]

Frandsen, Brigham R. 2017. "Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete." In *Regression Discontinuity Designs,* 38: 281–315. Emerald Group Publishing Limited. [81, 106]

Fuchs, Hans-Werner, and Lutz R Reuter. 1995. *Bildungspolitik seit der Wende: Dokumente zum Umbau des ostdeutschen Bildungssystems (1989-1994).* Opladen: Leske+ Budrich. [76, 112]

Gamoran, Adam, and Robert D Mare. 1989. "Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?" *American Journal of Sociology* 94 (5): 1146–83. [75]

Gartner, Hermann, and Thomas Hinz. 2009. "Geschlechtsspezifische Lohnungleichheit in Betrieben, Berufen und Jobzellen (1993–2006)." *Berliner Journal für Soziologie* 19 (4): 557–75. [73]

Goldin, Claudia, and Lawrence F. Katz. 2016. "A most egalitarian profession: Pharmacy and the evolution of a family-friendly occupation." *Journal of Labor Economics* 34 (3): 705–46. [73]

Goldin, Claudia, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth. 2017. "The expanding gender earnings gap: Evidence from the LEHD-2000 Census." *American Economic Review* 107 (5): 110–14. [73]

Helbig, Marcel, and Rita Nikolai. 2015. "Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949," 383. [79, 115]

Horstschräer, Julia, and Grit Muehler. 2014. "School entrance recommendation: A question of age or development?" *Education Economics* 22 (3): 270–92. [78, 113]

Hövel, Werner Van den. 2009. "Beginn der Schulpflicht in den Ländern." *Recht der Jugend und des Bildungswesens* 57 (2): 260–61. [79, 114, 115]

Humlum, Maria Knoth, Anne Brink Nandrup, and Nina Smith. 2019. "Closing or reproducing the gender gap? Parental transmission, social norms and education choice." *Journal of Population Economics* 32 (2): 455–500. [75]

Keane, Michael P, and Kenneth I Wolpin. 1997. "The career decisions of young men." *Journal of Political Economy* 105 (3): 473–522. [75]

Kelly, Ira R., Dhaval M. Dave, Jody L. Sindelar, and William T. Gallo. 2014. "The impact of early occupational choice on health behaviors." *Review of Economics of the Household* 12 (4): 737–70. [99]

Kratzmann, Jens, and Thorsten Schneider. 2009. "Social inequality, child care attendance, and school start in Germany." *Journal of Contextual Economics–Schmollers Jahrbuch,* no. 2, 181–90. [82]

Levine, Adeline. 1976. "Educational and occupational choice: A synthesis of literature from sociology and psychology." *Journal of Consumer Research* 2: 276–89. [75]

Morrison, Michael, and Neal J. Roese. 2011. "Regrets of the typical american: Findings from a nationally representative sample." *Social Psychological and Personality Science* 2 (6): 576–83. [99]

Mühlenweg, Andrea M., and Patrick A. Puhani. 2010. "The evolution of the school-entry age effect in a school tracking system." *Journal of Human Resources* 45 (2): 407–38. [92]

Muller, Daniel, and Lionel Page. 2016. "Born leaders: Political selection and the relative age effect in the Us congress." *Journal of the Royal Statistical Society Series a: Statistics in Society* 179 (3): 809–29. [74]

Oakes, Jeannie. 2005. *Keeping track: How schools structure inequality.* Yale University Press. [75]

Page, Lionel, Dipanwita Sarkar, and Juliana Silva-Goncalves. 2017. "The older the bolder: Does relative age among peers influence children's preference for competition?" *Journal of Economic Psychology* 63: 43–81. [74]

Page, Lionel, Dipanwita Sarkar, and Juliana Silva-Goncalves. 2019. "Long-lasting effects of relative age at school." *Journal of Economic Behavior & Organization* 168: 166–95. [74, 97]

Penner, Andrew M. 2008. "Race and gender differences in wages: The role of occupational sorting at the point of hire." *Sociological Quarterly* 49 (3): 597–614. [73]

Philipp, Julia. 2023. "Gendered university major choice: The role of intergenerational transmission." *Journal of Population Economics* 36 (2): 1049–97. [75]

Ramaci, Tiziana, Monica Pellerone, Caterina Ledda, Giuseppe Presti, Valentina Squatrito, and Vincenzo Rapisarda. 2017. "Gender stereotypes in occupational choice: A cross-sectional study on a group of Italian adolescents." *Psychology Research and Behavior Management* 10: 109–17. [98]

RDC of the Federal Statistical Office and Statistical Offices of the Federal States of Germany. 2014-2019. *Microcensus.* https://doi.org/10.21242/12211.2014.00.00.1.2.1, https://doi.org/10.21242/12211.2015.00.00.1.2.1, https://doi.org/10.21242/12211.2016.00.00.1.2.1, https://doi.org/10.21242/12211.2017.00.00.1.2.1, https://doi.org/10.21242/12211.2018.00.00.1.2.3, https://doi.org/10.21242/12211.2019.00.00.1.2.1.

Slavin, Robert E. 1990. "Achievement effects of ability grouping in secondary schools: A best-evidence synthesis." *Review of Educational Research* 60 (3): 471–99. [75]

Thompson, Angus H, Roger H Barnsley, and James Battle. 2004. "The relative age effect and the development of self-esteem." *Educational Research* 46 (3): 313–20. [74]

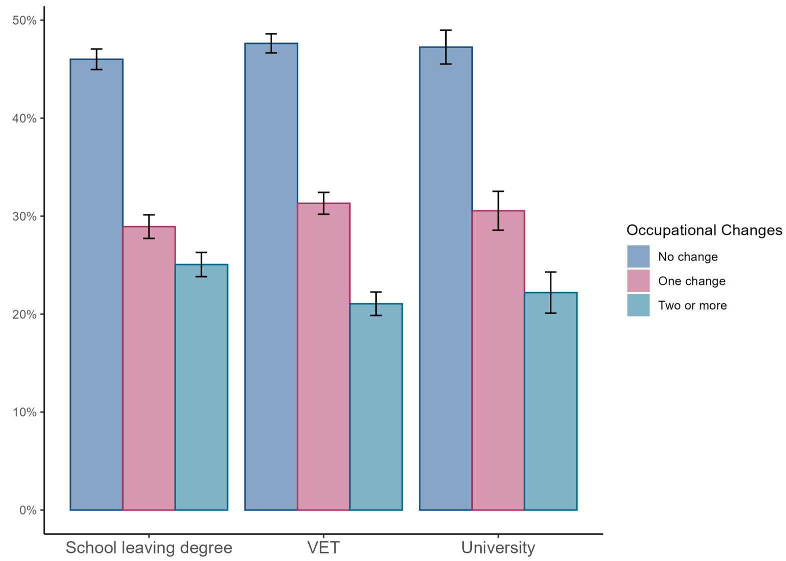
Vattuone, Giulia. 2026. "Worker sorting and the gender wage gap." Unpublished Manuscript. [75]

Vleuten, Maaike van der, Eva Jaspers, Ineke Maas, and Tanja van der Lippe. 2018. "Intergenerational transmission of gender segregation: How parents' occupational field affects gender differences in field of study choices." *British Educational Research Journal* 44 (2): 294–318. [75]

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the workplace, investment in human capital, and gender." *Quarterly Journal of Economics* 133 (1): 457–507. [75]

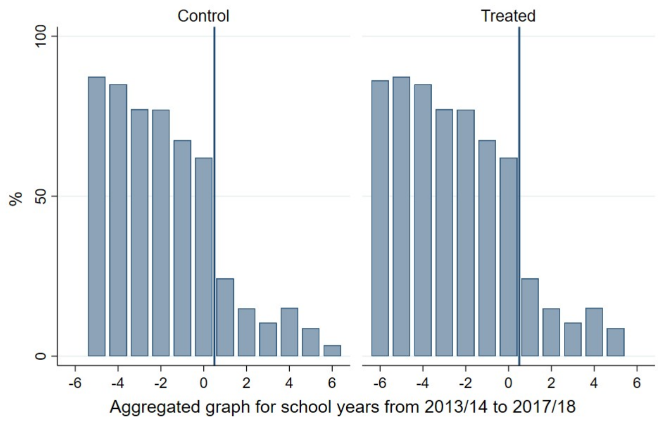
Zafar, Basit. 2013. "College major choice and the gender gap." *Journal of Human Resources* 48 (3): 545–95. [75]

Appendix 2.A Additional tables and figures



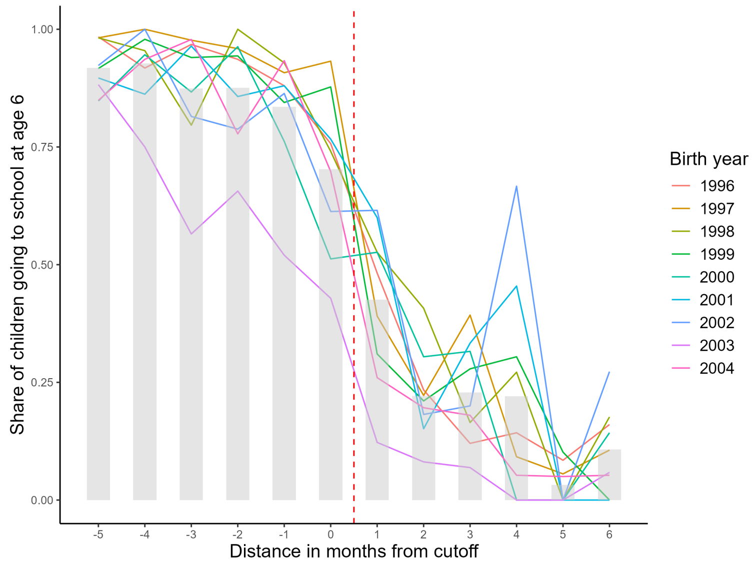
Notes: This figure shows the share of individuals changing occupations by highest educational level achieved.
Source: SOEP v37, own calculations. N = 46,941

Figure 2.A.1. Changes in occupation by highest educational level



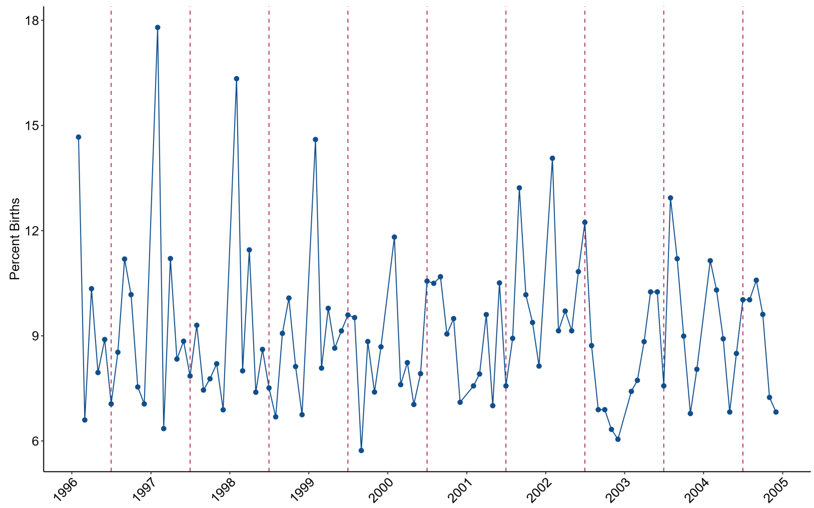
Notes: The figure plots the compliance rate with the school entry cutoff rules for both control and treated states, for the school years from 2013/2014 to 2017/2018. On the x axis, month of birth is normalized as to be equal to 0 for the last month before the cutoff date, and is equal to 1 for the first month after the cutoff. That is, in the control group, for example, the cutoff is June 30. This means that children born in June have *cutoff* = 0, and those born in July *cutoff* = 1, and so on. Source: Microcensus 2014-2019, own calculations. N = 3,134

Figure 2.A.2. Complience rate for school years 2013/2014 to 2017/2018



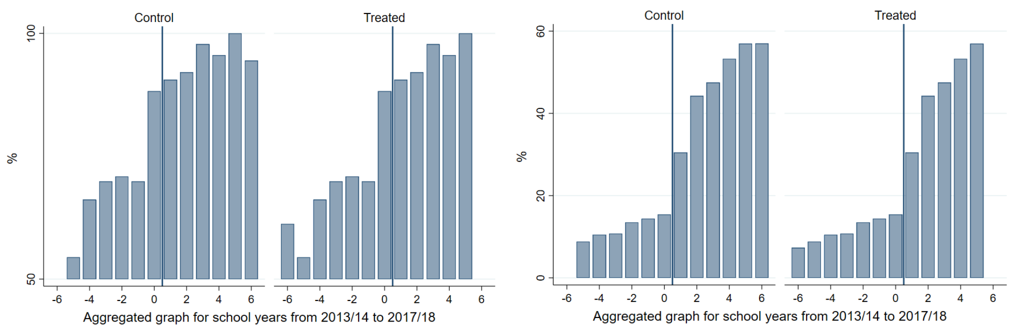
Notes: The figure plots the share of children going to school at age six for each cohort. On the x axis, month of birth is normalized as to be equal to 0 for the last month before the cutoff date, and is equal to 1 for the first month after the cutoff. That is, in the never treated states, for example, the cutoff is June 30. This means that children born in June have *cutoff* = 0, and those born in July *cutoff* = 1, and so on. Source: SOEP v37, own calculations. N = 3,088

Figure 2.A.3. Share of children going to school at age six



Notes: The figure plots the share of births by month (our running variable) in the relevant cohorts from 1996 to 2004, normalized by distance to the cutoff. When cutoff is 0 or lower, individuals should go to school when they are aged six. Dotted red lines show where school cutoff months is equal to 0. If parents had strong preferences for their children to be born either before or after the cutoff, we should see significant changes in trends around the cutoff. While some evidence of seasonality is shown here, it seems to be orthogonal to school cutoff dates. Source: SOEP v37, own calculations. N = 18,320

Figure 2.A.4. Patterns in births by month for relevant cohorts



Panel A: Compliance for low track secondary school

Panel B: Compliance for high track secondary school

Notes: In this figure, Panel A plots the probability of being still in grade nine, conditional on attending a low track secondary school (N = 6,965), while in panel B the probability of being still in grade ten, conditional on being in the high track secondary school (N = 4,310). The plots are for both control and treated states for the school years 2013/2014 to 2017/2018. On the x axis, month of birth is normalized as to be equal to 0 for the last month before the cutoff date, and is equal to 1 for the first month after the cutoff. That is, in the control group, for example, the cutoff is June 30. This means that children born in June have *cutoff* $= 0$, and those born in July *cutoff* $= 1$, and so on. So, for example, for the high track students, the probability of being still in grade ten increases by 50% (15 p.p.) for those born after the cutoff in the treated states.
Source: Microcensus 2014-2019, own calculations.

Figure 2.A.5. Discontinuities in grades by secondary school track

Table 2.A.2. Results of manipulation test (p-values) at various cutoffs

	P-values at different K		
	K = 0	**K = 0.01**	**K = 0.05**
Panel A: Relevant Cohorts (n = 18,320)			
Pooled	0.024**	0.031**	0.273
Treated (n = 507)	0.333	0.336	0.383
Control (n = 17,813)	0.009***	0.013**	0.161
Pooled, placebo for August	0.775	0.788	0.936
Panel B: Pre-Reform Cohorts (n = 37,483)			
Pooled	0.000***	0.000***	0.000***
Placebo, February	0.000***	0.000***	0.000***
Placebo, April	0.001***	0.003***	0.191
Placebo, September	0.107	0.142	0.719
Placebo, November	0.000***	0.000***	0.000***

Notes: This table presents p-values from the manipulation test around the reform cutoffs proposed by Frandsen (2017). The test is designed for regression discontinuities where the running variable is discrete, as in our case (month of birth). Specifically, we test the null hypothesis of absence of manipulation by assessing whether the distribution of births around the cutoff deviates from a linear trend, meaning there is no unusual heaping. We present the results for our relevant cohorts (birth years 1996 to 2004) as well as for pre-reforms cohorts (birth years 1989 to 1995) to check whether these trends changed over time. The parameter K used in the test represents how much nonlinearity is allowed in the distribution of births around the threshold. When $K = 0$, the test is is likely to reject the null even if there is a small nonlinearity. Low p-values (< 0.05) thus suggest evidence of manipulation. For example, significant evidence of manipulation is found for the relevant cohorts at lower values of K, while the treated group shows no manipulation. The pre-reform cohorts display strong evidence of manipulation, with consistently low p-values. Results from placebo tests are also reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SOEP v37, own calculations.

Table 2.A.1. Effect of treatment on school enrollment: Robustness to bandwidth choice

	Treatment Effect		
	LATE	Half-BW	Double-BW
Panel A: Bandwidth = 4			
Cutoff (not controlling for gender)	-0.161**	-0.122**	-0.280***
	(0.072)	(0.062)	(0.051)
Cutoff (controlling for gender)	-0.155**	-0.111*	-0.276***
	(0.078)	(0.066)	(0.056)
Mean Control	0.255	0.318	0.212
Observations	1889	758	3147
Panel B: Bandwidth = 5			
Cutoff (not controlling for gender)	-0.205***	-0.054	-0.295***
	(0.061)	(0.094)	(0.050)
Cutoff (controlling for gender)	-0.198***	-0.047	-0.291***
	(0.067)	(0.101)	(0.054)
Mean Control	0.224	0.318	0.212
Observations	2393	1318	3165
Panel C: Bandwidth = 6			
Cutoff (not controlling for gender)	-0.232***	-0.057	-0.303***
	(0.054)	(0.094)	(0.049)
Cutoff (controlling for gender)	-0.226***	-0.050	-0.300***
	(0.059)	(0.102)	(0.053)
Mean Control	0.212	0.283	0.212
Observations	2838	1318	3167

Notes: This table presents regression discontinuity estimates of distance from school entry dates cutoff on school enrollment using different bandwidth choices. The variable *Cutoff* is the normalized distance from the last month before the cutoff date (equal to 0). So, for example, in never treated states, the cutoff is June 30. This means that children born in June have cutoff = 0, and those born in July cutoff = 1. Each panel shows results for a different bandwidth (4, 5, and 6 months), with estimates using the main bandwidth (LATE), half the bandwidth, and double the bandwidth. For each specification, we present estimates both with and without controlling for gender. The *Mean Control* row shows the probability of starting school at age six for observations just above the cutoff within each bandwidth. Robust standard errors clustered at the state level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SOEP v37, own calculations.

Table 2.A.3. Balance table of sample characteristics between treatment and control states

	Control States	Treated states	Mean difference
	mean/sd	mean/sd	
	(1)	(2)	(3)
Gender	0.486	0.487	-0.000703
	(0.500)	(0.500)	(0.00222)
Age	18.30	18.39	-0.0098***
	(2.387)	(2.397)	(0.0106)
Had already worked	0.0925	0.0959	-0.00337**
	(0.290)	(0.294)	(0.00164)
Works full time	0.731	0.751	-0.0200***
	(0.443)	(0.432)	(0.00320)
German citizenship	0.949	0.943	0.00617***
	(0.326)	(0.348)	(0.00147)
Completed VET	0.0926	0.108	-0.0153***
	(0.290)	(0.310)	(0.00131)
Abitur	0.367	0.360	0.00698**
	(0.482)	(0.480)	(0.00291)
University degree	0.00957	0.0117	-0.00213***
	(0.0974)	(0.108)	(0.000443)
Nonmissing ISCO3 code	0.297	0.319	-0.0215***
	(0.457)	(0.466)	(0.00204)
% women in occupation	0.463	0.452	0.0107***
	(0.312)	(0.313)	(0.00253)
25% segr. occupation	0.599	0.607	-0.00755*
	(0.490)	(0.489)	(0.00396)
10% segr. occupation	0.214	0.224	-0.0100***
	(0.410)	(0.417)	(0.00333)
N	65,550	222,884	288,434

Notes: This table compares sample characteristics between individuals in treatment and control states. The table presents means and standard deviations for various demographic and occupational variables, including gender, age, work status, citizenship, education levels, and occupational segregation measures. Column (3) shows the mean differences between treated and control states. Microcensus 2014-19, own calculations.

Table 2.A.4. Robustness check: RDD with different Fixed Effects specifications on school entry rules compliance

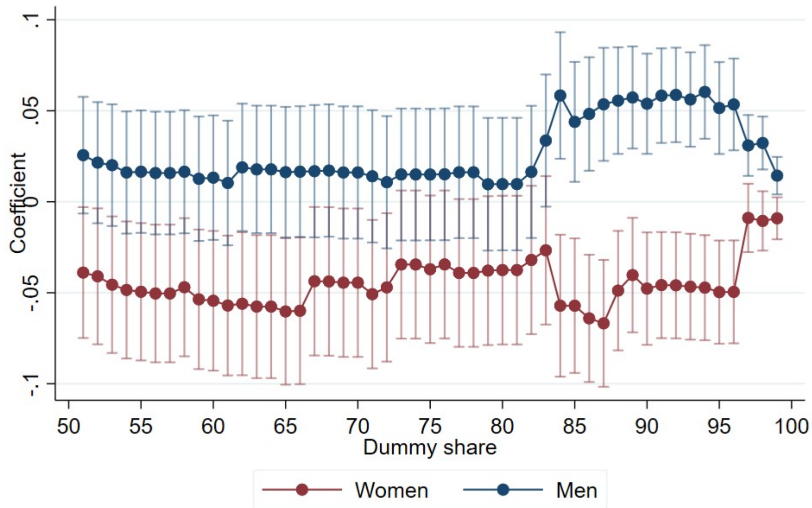
	Starting school at 6			Grade 9			VET			Grade 10 (Gymnasium)			Grade 11 (Gymnasium)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Before	0.64***	0.40***	0.40***	−0.29***	−0.12**	−0.14**	0.031**	0.012*	0.014*	−0.34***	−0.17***	−0.17***	0.13***	0.094	0.095
	(0.034)	(0.088)	(0.088)	(0.079)	(0.042)	(0.048)	(0.012)	(0.0069)	(0.0072)	(0.021)	(0.043)	(0.034)	(0.030)	(0.11)	(0.085)
Male	−0.017	−0.018	−0.018	−0.028	−0.023	−0.025	0.0027	0.0029	0.0027	0.0080	0.018	0.013	−0.058**	−0.068***	−0.062**
	(0.016)	(0.015)	(0.016)	(0.026)	(0.022)	(0.026)	(0.0022)	(0.0020)	(0.0021)	(0.012)	(0.013)	(0.012)	(0.023)	(0.022)	(0.021)
Before × Male	−0.036*	−0.035*	−0.034*	0.044	0.029	0.043	0.0052	0.0065	0.0054	−0.00068	−0.012	−0.0061	−0.051*	−0.040	−0.047*
	(0.018)	(0.017)	(0.017)	(0.030)	(0.036)	(0.031)	(0.0048)	(0.0047)	(0.0048)	(0.020)	(0.019)	(0.018)	(0.028)	(0.025)	(0.025)
February			0.0087			0.098			−0.024			0.012			−0.044*
			(0.024)			(0.10)			(0.014)			(0.019)			(0.023)
March			−0.014			0.11*			−0.014			0.034**			0.038
			(0.031)			(0.060)			(0.012)			(0.014)			(0.030)
April			−0.028			0.18			−0.023*			0.065***			−0.042
			(0.028)			(0.010)			(0.012)			(0.019)			(0.026)
May			−0.051**			0.13*			−0.020*			0.054***			−0.066**
			(0.027)			(0.065)			(0.011)			(0.018)			(0.028)
June			−0.097***			0.19**			−0.030**			0.055**			−0.034
			(0.026)			(0.087)			(0.013)			(0.021)			(0.032)
July			−0.094**			0.25***			−0.033**			0.045			0.029
			(0.043)			(0.080)			(0.012)			(0.031)			(0.075)
August			−0.21***			0.29**			−0.038**			0.16***			−0.014
			(0.055)			(0.10)			(0.013)			(0.024)			(0.086)
September			−0.28***			0.27***			−0.039***			0.21***			−0.021
			(0.058)			(0.088)			(0.013)			(0.040)			(0.089)
October			−0.39**			0.26**			−0.038***			0.29***			−0.040
			(0.14)			(0.093)			(0.013)			(0.048)			(0.10)
November			−0.43***			0.33***			−0.042***			0.31***			−0.12
			(0.12)			(0.10)			(0.013)			(0.031)			(0.087)
December			−0.43***			0.33***			−0.042***			0.31***			−0.16
			(0.088)			(0.094)			(0.013)			(0.057)			(0.10)
YOB FE	✓	–	✓	✓	–	✓	✓	–	✓	✓	–	✓	✓	–	✓
YOB×MOB FE	–	✓	–	–	✓	–	–	✓	–	–	✓	–	–	✓	–
MOB FE	–	–	✓	–	–	✓	–	–	✓	✓	–	✓	✓	–	✓
N	5,535	5,535	5,535	763	763	763	7,123	7,123	7,123	4,310	4,310	4,310	4,310	4,310	4,310
Adjusted R^2	0.349	0.400	0.400	0.085	0.112	0.094	0.012	0.020	0.014	0.149	0.170	0.170	0.018	0.026	0.024
F Statistic	157.4	15.4	1,099,300.5	4.90	4.99	100.3	5.16	2.56	83.4	170.3	28.5	539.0	12.2	15.1	69.9

Notes: This table presents the following results: Columns (1)-(3) show the probability change in starting school at age six when born before the cutoff for the cohorts 2007-2011. Columns (4)-(6) estimate the probability of being still in Grade nine for the cohorts 1998-2002, conditional on being enrolled in a low track secondary school (Hauptschule). Columns (7)-(9) show the probability of being already in VET, conditional on having obtained a school leaving degree from a low track secondary school for the cohorts 1998-2002. Columns (10)-(12) and (13)-(15), show, respectively, the probability of being still Grade ten or already in Grade eleven for the cohorts 1997-2001, conditional on being enrolled in a high track secondary school (Gymnasium). Columns (1),(4), (7), (10) and (13) are estimated with Year of Birth FE, columns (2), (5), (8), (11) and (14) have Year of Birth × Month of Birth FE; columns (3), (6), (9), (12) and (15) have Year and Month of Birth FE. In each case, the baseline changes and so the interpretation of the coefficients of interest. Robust standard errors are clustered at the state level. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.5. Main RDD model with different Fixed Effects specifications

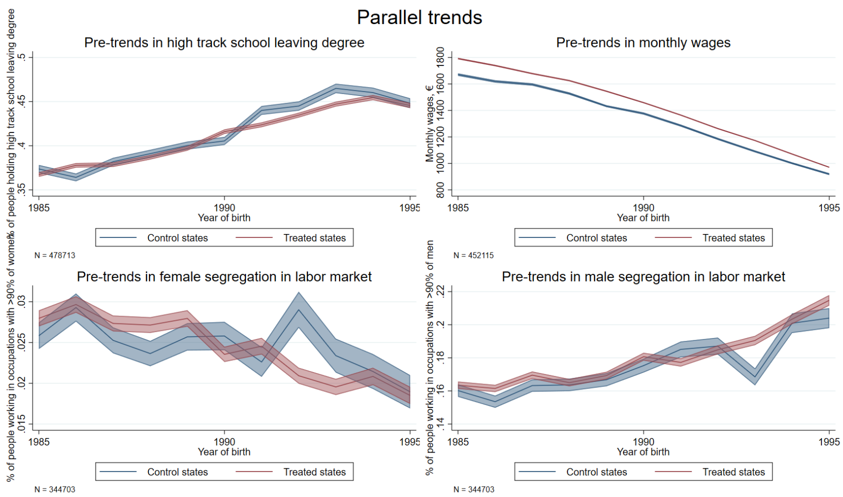
	Segr. occup. (25%)		Segr. occup. (10%)		Segr. occup. (25%)		Segr. occup. (10%)		Segr. occup. (25%)		Segr. occup. (10%)	
	All	Low track	All	Low track	All	Low track	All	Low track	All	Low track	All	Low track
Before	−0.014	−0.016	−0.0054	−0.0034	−0.023	−0.044*	0.018	0.010	−0.015	−0.026	0.0054	−0.0068
	(0.010)	(0.012)	(0.0050)	(0.0079)	(0.017)	(0.021)	(0.024)	(0.023)	(0.012)	(0.019)	(0.013)	(0.017)
Male	0.18***	0.19***	0.30***	0.36***	0.18***	0.19***	0.30***	0.36***	0.18***	0.19***	0.30***	0.36***
	(0.012)	(0.014)	(0.019)	(0.014)	(0.012)	(0.014)	(0.018)	(0.014)	(0.012)	(0.014)	(0.019)	(0.014)
Before × Male	0.044**	0.056**	0.025*	0.030**	0.045***	0.057**	0.026*	0.030**	0.044**	0.056**	0.026*	0.030**
	(0.015)	(0.021)	(0.013)	(0.014)	(0.015)	(0.021)	(0.013)	(0.014)	(0.015)	(0.021)	(0.013)	(0.014)
June									−0.010	−0.012	−0.017	−0.018
									(0.013)	(0.015)	(0.013)	(0.017)
July									−0.0090	−0.025	−0.0033	−0.021
									(0.019)	(0.025)	(0.019)	(0.026)
August									−0.0069	−0.018	0.011	−0.0024
									(0.022)	(0.027)	(0.022)	(0.032)
September									0.013	0.013	−0.021	−0.039
									(0.026)	(0.031)	(0.020)	(0.023)
October									0.0057	−0.010	0.022	−0.018
									(0.052)	(0.029)	(0.032)	(0.030)
November									−0.033	−0.053	0.075	0.044
									(0.050)	(0.050)	(0.055)	(0.046)
December									0.011	−0.047	−0.087	−0.13**
									(0.031)	(0.028)	(0.053)	(0.044)
YOB FE	✓	✓	✓	✓	–	–	–	–	✓	✓	✓	✓
YOB×MOB FE	–	–	–	–	✓	✓	✓	✓	–	–	–	–
MOB FE	–	–	–	–	–	–	–	–	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	17,294	11,958	17,294	11,958	17,294	11,958	17,294	11,958	17,294	11,958	17,294	11,958
Adj. R^2	0.045	0.054	0.136	0.157	0.044	0.054	0.137	0.158	0.045	0.054	0.137	0.157

Notes: This table presents the following results: Columns (1)-(4) show the regression of the effect of being born before the school entry cutoff date on the probability of working in a segregated occupation, controlling for year of birth. In columns (5)-(8), the same regression is presented with year × month of birth FE; and in columns (9)-(12): same regression is shown, but with year and month of birth FE. Throughout all specifications, we limit the sample to those born in the two months before and after the cutoff, so that the baseline is May for never treated states, and we run the regression first on all sample, then on the subsample of individuals with a low track school leaving degree (our main sample of interest). Further controls include migration background, ZIP code area (NUTS 3), interview quarter year. Robust standard errors are clustered at the state level. Microcensus 2014-19, own calculations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



Notes: The figure above shows the coefficients of the variable of interest D_i of the DiD specification, β_1 for women and β_3 for men, when choosing different thresholds for the indicator of occupational segregation. The results are stable across the different thresholds above which we define an occupation to be segregated or not.
Source: Microcensus 2014-2019, own calculations. N = 21,193

Figure 2.A.6. Coefficients' graph for different segregation thresholds



Notes: The figure shows the trends in relevant education and labor market outcomes in treated and control states for the cohorts born just before the school entry reforms (1985-1995). In detail, we see in the first quadrant the share of people holding the highest school leaving degree ('Abitur'), in the second the trends in monthly wages, and in the third and fourth the trends in female and male-dominated occupations. Due to data availability, we cannot show directly trends of the pre-period, but only current trends for individuals born in the pre-period, since we do not have repeated observations over time for the same individuals.
Source: Microcensus 2014-2019, own calculations.

Figure 2.A.7. Parallel trends of cohorts born before the reforms in treated and control states

Appendix 2.B Additional institutional background

2.B.1 The German education system

In Germany, education and thus also the educational system falls under the authority of the 16 federal states. To coordinate certain standards, however, there exists the assembly of ministers of education of the German states called "Kultusministerkonferenz" (KMK) that regularly decides on non-binding directives for harmonization and common development in the states. Thus, while the system of schooling varies in its details across states, there are important common components. Further, the educational system underlies constant changes. In the paper, we mention the common components across time and states. We only highlight changes and differences in case they matter for our identification strategy. It is important to note that the states that were part of the GDR displayed a markedly different system between 1946 and 1990. In particular, the school system of the GDR did not contain early tracking. After the reunification with the school year 1991/92, the school system of the states of the former GDR was harmonized towards the system of the West German states(see e.g., Fuchs and Reuter, 1995, for details). In Saxony, this was delayed to the school year 1992/93 (Fuchs and Reuter, 1995).

2.B.2 School entry age in Germany

The beginning of compulsory schooling, and thus the age of school entrance, is subject to state legislation. However, ever since the Hamburg Agreement from 1964, this was harmonized across West German states and only recently diverged again. Prior to the recent reform wave, children had to enter primary school in the year in which they are six years old by June 30. This cutoff goes back to Prussian regulations from 1922 and was put into law in the whole German Reich in 1938 (see Reichsschulpflichtgesetz, 1938). With the school year 1941, the cutoff was moved to December 31. After World War II, some differences emerged across states. In West Germany, these were unified after the Hamburg Agreement of the KMK, 1964, which suggested shifting the cutoff uniformly back to June 30 (see Faust, 2006).

In East Germany, the cutoff was May 31 prior to reunification (see Gesetz über die sozialistische Entwicklung des Schulwesens in der Deutschen Demokratischen Republik, 2.12.1959). In the process of harmonizing the GDR school system toward the West German system, the school entry cutoff was also changed to June 30. In the majority of East German states, this was implemented from the school year 1991/92 onwards. Only in Thuringia, the new cutoff was introduced in the school year 1993/94. In the preliminary law of Thuringia from March 25, 1991, it states that all children must enter primary school when they are six years old at the beginning of the school year. The beginning of the school year, however, is not specified in the law (see Gesetz- und Verordnungsblatt (GVBl.) für den Freistaat Thüringen, No. 3, 1991, p. 43). The cutoff of June 30 was introduced with the final education law of

August 6, 1993 (see GVBl für den Freistaat Thüringen, No. 21, 1993, p. 445). That is, the cutoff was introduced in the school year 1993/94. It is unclear which cutoff applied in the school years 1991/92 and 1992/93.

In Saxony, the cutoff of June 30 was introduced with the education law of July 1, 1991 (see SächsGVBl., No. 13, 1991, p. 213). Thus, it applied from the school year 1991/92 onward. In Mecklenburg-Western Pomerania, the cutoff was introduced with the education law of April 26, 1991 in the school year 1991/92 (see Erstes Schulreformgesetz des Landes Mecklenburg-Vorpommern (SRG), April 26, 1991, GVBl., No. 8, 1991, p. 125). In Saxony-Anhalt, it was introduced with the education law of July 11, 1991 with the school year 1991/92 (see Schulreformgesetz für das Land Sachsen-Anhalt vom July 11, 1991, GVBl., No. 17, 1991, p. 165).

In 1997, the German states agreed in a KMK to allow for more flexibility in the cutoff with the aim to make students enter primary school earlier. Eight of the sixteen federal states moved the month-of-birth cutoff away from June 30. This process differed across states, as it did not start simultaneously, took longer in some states, the new cutoff dates differed, and the distance between the old and new cutoff dates was not the same in all states. In the cases of Berlin, Bavaria, and Lower Saxony, the cutoff dates were reverted back to previous cutoffs in the 2010s and 2020s. However, as we will not observe the affected cohorts in our sample yet, we will ignore these reversals for now.

2.B.2.1 Early and delayed entry

In most German states, parents can apply to delay their child's entry if they believe the child is not ready for school, usually due to developmental concerns or lack of school readiness ("Schulfähigkeit"). The process typically requires an application, sometimes accompanied by a pediatrician's or psychologist's assessment, and must be submitted several months before the school year starts. In some cases, the school management decides based on medical advice and a school readiness assessment. In practice, postponement is relatively accessible, especially for children born close to the cutoff date, but it does require formal steps and sometimes professional input (Horstschräer and Muehler, 2014).

Early school entry is possible but generally less common and subject to stricter requirements. Children usually must demonstrate advanced physical, intellectual, and social development according to a pediatrician or psychologist assessment, and often school principals have the final say in whether to admit the child one year earlier (Horstschräer and Muehler, 2014).

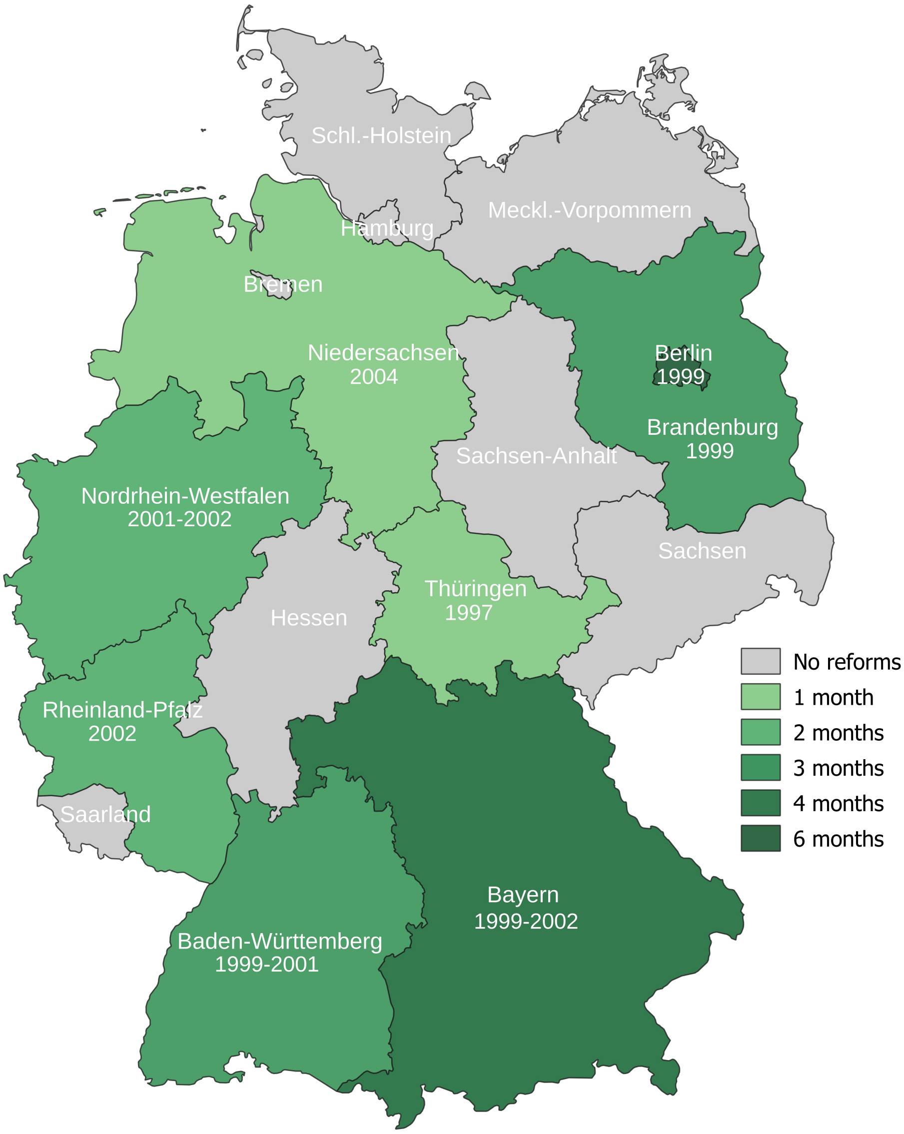
2.B.2.2 Early entry

In Baden-Württemberg, Saxony-Anhalt, Schleswig-Holstein, and Thuringia early entry upon 'Schulfähigkeit' is only possible for children who turn six by June 30 of the following year; in Berlin that date is March 31; in Bremen, Brandenburg, and

Mecklenburg-Western Pomerania, children can enter early if they turn six before December 31 of that year. In Saxony, children that turn six years old between June 30 (normal cutoff) and September 30 can—but do not have to—go to school in that year; children that turn six after September 30 can enter early if they display 'Schulfähigkeit'. The dates reflect rules as of November 2008 and are taken from Hövel (2009). Exact dates may be subject to changes over time.

2.B.2.3 Delayed entry

With respect to delayed entry, there are some states that have stricter or more lenient rules. In Berlin and Schleswig-Holstein, there are no provisions for delay; in Rhineland-Palatinate and North Rhine-Westphalia, there need to be severe or important health reasons to allow a delay; in Saxony-Anhalt, delay is decided on an individual case basis. Rules are based on November 2008 and taken from Hövel (2009). Exact rules may be subject to changes over time.



Notes: Map of school entry cutoff date reforms in Germany for the cohorts born between 1997 and 2004. Different colors depict the 'intensity' of the reforms, i.e., how many months the cutoff was shifted by each federal state from the baseline, June 30. Sources: Faust (2006), Berthold (2008), Hövel (2009), and Helbig and Nikolai (2015).

Figure 2.B.1. School entry cutoff date reforms in Germany, cohorts 1997-2004

Chapter 3

Does Stress Impact Risk-Taking?*

Joint with Si Chen, Thomas Dohmen, Anna Schulze Tilling and Elena Shvartsman

3.1 Introduction

Risk-taking is fundamental to decision-making in domains such as finance, health, and social behavior, with significant personal and societal consequences. Yet, a key puzzle in economics is that the same individual sometimes makes different choices when faced with the identical risky decision, challenging the notion of stable preferences. One natural candidate for this variability is acute stress, which is pervasive in real-world decision contexts, ranging from financial markets to medical emergencies, and affects decision-critical regions of the brain (see Starcke and Brand, 2012, for a review), potentially shaping a wide range of behaviors. However, despite years of study, evidence on the impact of acute stress on risk-taking remains contradictory. Some studies find that stress increases risk seeking (e.g., Buckert et al., 2014; Benda-

* We would like to express our utmost gratitude to Holger Gerhardt and Simone Jost for their outstanding support in the data collection process. We are also grateful to Paul Behler, Timo Freyer, Ilona Krupp, and Malin Siemers for helping us collect the data. Excellent research support was provided by Johannes Beckers, Niklas Bongers, Matthew Esche, Max Jahn, Daniel Jung, Elisa Klemann, Lukas Kutsch, Justus Ratz, Jadine Rauh, Lennart Schultze, Yashaswini Shekhawat, Annalena Tetzner, Gereon tho Pesch, and Simon Wolfers. We are grateful for comments and suggestions by Jana Cahlíková, Bernadette von Dawans, Thomas Epper, Holger Gerhardt, Klemens Kirschbaum, John List, and Jan Scherer. We also thank audiences at the CERGE-EI, Fernuniversität Hagen, Universities of Bonn, Lüneburg, Mannheim, Milan, WHU – Otto Beisheim School of Management, and participants of the European ESA meeting 2023, IMEBESS 2024, MBEES 2024, TIBER 2023, NGE Workshop Hohenheim 2022, and Verein für Socialpolitik 2024 for helpful feedback and suggestions. Funding by the Diligentia Foundation for Empirical Research is gratefully acknowledged. Further funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A05), and under Germany's Excellence Strategy EXC 2126/1-390838866 is gratefully acknowledged. The study was approved by the Ethics Committee of the Faculty of Medicine at the University of Bonn (#021/19) on 02.04.2019 and 05.13.2022 (amendment). The study was preregistered on AsPredicted.org under #100204, https://aspredicted.org/t3pv74.pdf.

han et al., 2017), others report that it increases risk aversion (e.g., Yamakawa et al., 2016; Cahlíková and Cingl, 2017), and yet others find no effects (e.g., Sokol-Hessner et al., 2016; Veszteg et al., 2021; Parslow and Rose, 2022).[1] The inconsistencies in previous evidence may reflect both theoretical ambiguity and methodological limitations in the study of stress and risk-taking.

On the theoretical side, most studies implicitly assume a homogeneous effect of stress on risk-taking, i.e., that stress makes all decisions more risk-seeking or more risk-averse. However, stress might operate more subtly. In particular, it could (i) push risk-averse decisions toward risk seeking and risk-seeking decisions toward risk aversion (a shift toward neutrality), or (ii) amplify existing tendencies, making risk-averse decisions more risk-averse and risk-seeking decisions more risk-seeking (a shift away from neutrality). Both patterns are plausible given the biological effects of acute stress on decision-critical brain regions.[2]

On the methodological side, most existing studies have limitations that complicate causal inference. First, stress is often not measured continuously during the decision-making phase, leaving uncertainty about whether participants were indeed under acute stress when their risk-taking behavior was measured.[3] Second, most studies rely on between-subject designs with relatively small samples.[4] Between-subject designs risk confounding the causal effect of stress with heterogeneity in participants' baseline risk preferences. Differences in the distribution of measured risk preferences between treatment and control groups, stemming either from true latent preferences (i.e., sampling variation) or from measurement error, can be mistakenly attributed to stress. These sampling and measurement issues offer a potential ex-

1. See Table 3.A.1 in Appendix 3.A for a detailed overview.

2. Specifically, several brain regions are involved in risky decision-making, with different regions specializing in distinct aspects. The striatum, a subcortical reward-related structure, encodes the expected value (i.e., the mean outcome) and is relatively insensitive to risk (Tobler et al., 2007; Tobler, Dolan, and Dayan, 2009). In contrast, the lateral orbitofrontal cortex (lateral OFC), a part of the prefrontal cortex, exhibits increased activity in response to increasing risk (variance), particularly in more risk-averse individuals. Conversely, in risk-seeking individuals, increased risk is associated with stronger activation in more medial prefrontal cortex regions (medial OFC) (Tobler et al., 2007). A shift toward or away from risk neutrality may depend on which pathway dominates.

3. Notable exceptions include Bendahan et al. (2017) and Cahlíková and Cingl (2017).

4. A small number of studies have employed within-subject designs to study the effect of acute stress on risk-taking, most notably Clark et al. (2012), Cueva et al. (2015), Daughters et al. (2013), Von Helverson and Rieskamp (2013) and Sokol-Hessner et al. (2016). While these papers represent important methodological advances, their analysis or effective statistical power is constrained by specific design features. For example, Daughters et al. (2013) collect control and stress observations on the same day and in a fixed order, and Sokol-Hessner et al. (2016) split the sample across multiple treatment sequences to account for time effects, reducing the number of observations available for the key control–stress comparison. Interestingly, the two most highly powered between-subject studies, Parslow and Rose (2022) and Veszteg et al. (2021), find no effect of stress on risk-taking. Our design builds on this literature by implementing a high-powered within-subject framework that cleanly isolates the causal effect of acute stress. By separating control and stress sessions across days, counterbalancing order, and focusing directly on the control–stress contrast, we increase statistical power while avoiding confounds related to order effects, time trends, and stress-induction validity.

planation for the contradictory findings in the literature. Albeit, the continued use of small-sample designs is understandable given the costly and resource-intensive nature of inducing and measuring acute stress.

We develop a novel experimental design that addresses the theoretical and methodological issues outlined above while substantially increasing statistical power. We employ a within-subject design in which each participant came to the laboratory twice and completed the same set of incentivized risky choices: once in a *Control* and once in a *Stress* condition. This design controls for baseline heterogeneity in risk preferences, as each participant serves as their own control, i.e., baseline risk preferences are by construction identical across control and treatment groups. It also increases statistical power, enabling us to detect both homogeneous and heterogeneous shifts in risk-taking induced by acute stress. Importantly, these gains in terms of statistical power come at a relatively low cost: as noted by List (2025), within-subject designs can achieve acceptable statistical power with fewer participants than comparable between-subject designs in certain settings.

We induce acute stress using a modified version of the well-established Trier Social Stress Test for Groups (TSST-G, Von Dawans, Kirschbaum, and Heinrichs, 2011). In contrast to many studies that expose participants to a brief stressor before the main task, we embed the risky choices within the stress procedure itself, ensuring that participants remain stressed throughout all risky decisions. Moreover, stress levels were continuously monitored using biomarkers (salivary cortisol and heart rate variability) as well as participants' self-reports.[5]

We do not find evidence of a significant causal effect of acute stress on risk-taking. Specifically, we neither find evidence of a homogeneous shift toward more risk-seeking or risk-averse behavior, nor a shift toward or away from risk-neutrality. Notably, a between-subject analysis based on data from participants' first session, when half of our participants were assigned to the *Control* and half to the *Stress* condition, suggests that stress reduces risk aversion. However, our better-powered within-subject analysis does not support such a causal effect, nor does it indicate shifts toward or away from risk neutrality. These results suggest that the apparent effects in the between-subject analysis are likely driven entirely by differences in baseline risk preferences between treatment and control groups. Given that our sample is larger than most previous studies,[6] our procedures in eliciting risk-taking are state-of-the-art, and our rigorous collection of biomarker data confirms participants were in a state of acute stress throughout all risky decisions in the *Stress* condition, it

5. The stress protocol of TSST-G features a mock job interview, which resembles a real-life situation and, hence, induces a stress level comparable to challenging daily situations. Our approach combines continuous stress induction with continuous measurement, allowing us to closely track participants' stress levels during decision-making.

6. After applying our preregistration criteria, we have a main sample of N=142, and a total of 6,805 decisions, which is substantially more than most related studies. Compare Table 3.A.1 in Appendix 3.A for an overview of related literature and Table 3.B.1 for power calculations.

is unlikely that the divergence between our between-subject results and our better-powered within-subject results is specific to our experimental setting. Instead, our between-subject results are representative of the literature: much of the variation in findings can possibly be attributed to differences in risk preferences between treatment and control groups, due to sampling variation.

Our results provide a foundational step toward evaluating whether acute stress shapes risk-taking behavior. Our well-powered null results boost our confidence that acute stress does not affect risk-taking,[7] especially in settings involving small-stakes financial choices. Our null result thereby provides useful practical implications, for instance, for the delegation of tasks and the allocation of decision rights. If employees can handle small-stakes financial decisions effectively even when stressed, organizations do not need to worry about the impact of stress, pervasive in everyday working environments, when delegating such decisions to employees.

This study makes two primary contributions to the literature. First, substantively, we do not find any evidence that acute stress, on average, shifts individuals toward increased or decreased risk-taking. The strength of this result relies on our extraordinarily clean design: since we continuously induced acute stress throughout the decision phase and also continuously tracked the stress induction, we can rule out that our results are driven by methodological flaws in the stress induction. Further, our within-subject design and corresponding analysis provide us with greater statistical power than conventional between-subject designs (see also List, 2025). Our results help reconcile prior contradictory findings: in between-subject studies, apparent shifts in risk-taking can arise from baseline heterogeneity between participants, which may create spurious increases or decreases depending on the distribution of preferences in the sample.

Second, our analysis explicitly tests for heterogeneity in treatment effects. We are the first to examine whether acute stress shifts risk-taking toward or away from risk neutrality. Such a pattern could also have reconciled previous findings and would constitute an economically meaningful effect. We find no evidence of such a pattern, further strengthening the interpretation of our null results and increasing our confidence that there is no causal effect of acute stress on risk-taking, particularly in small-stakes financial choices.

This paper proceeds as follows: Section 3.2 describes our experimental design, and Section 3.3 outlines our analysis. Finally, Section 3.4 discusses policy implications and concludes.

7. In Section 3.4, we discuss how our null results update the weight of evidence relative to previous studies (see also List, 2026).

3.2 Experimental design and data

3.2.1 Overview and chronological order of the experiment

We conducted a laboratory experiment in which participants attended two sessions held one week apart, completing one session under acute stress (*Stress* condition) and one without exogenously induced stress (*Control* condition). To induce acute stress, we used a modified version of the Trier Social Stress Test for Groups (TSST-G, Von Dawans, Kirschbaum, and Heinrichs, 2011). The *Control* condition mirrored all procedural aspects of the *Stress* condition without the stress-inducing components.[8] The order of the conditions was counterbalanced: 49% of participants experienced the *Stress* condition first, while 51% completed the *Control* condition first.

In both conditions, we elicited participants' certainty equivalents for the same 28 lotteries using a multiple price list (MPL) format. The lotteries were taken from Bruhin, Fehr-Duda, and Epper (2010). Specifically, we take the same subset that they based their Beijing 2005 experiment on. The order of the lotteries was randomized across individuals and sessions.

To verify the effectiveness of the stress induction, we repeatedly collected salivary cortisol samples and continuously monitored participants' heart rates throughout each session. In addition to these physiological measures, participants rated their subjective stress levels. We also assessed participants' chronic stress using the validated Trier Inventory for Chronic Stress (TICS, Schulz and Schlotz, 1999) and hair cortisol concentration (HCC, Stalder and Kirschbaum, 2012).

Figure 3.1 illustrates the timeline of a session. Each session began with an one-hour adaptation phase designed to eliminate potential pre-treatment stress or arousal (cf. Dohmen, Rohde, and Stolp, 2023). This adaptation was followed by the treatment phase, which included three consecutive components, referred to as 'challenges' in the *Stress* condition and 'tasks' in the *Control* condition. These components differed in their stress-inducing properties, which we detail below.

After the first challenge/task, participants completed the first half of the 28 lottery decisions (14 decisions). They completed the second half after the second challenge/task. They then faced the third challenge/task, immediately after which we recorded participants' self-assessed stress levels. Subsequently, participants completed a series of questionnaires covering demographics, health, and lifestyle information relevant to cortisol measures and chronic stress. A full list of questionnaires is available in Appendix 3.C.9. At the end of the participants' second session, participants were invited to provide a hair sample for additional monetary compensation.

8. The TSST-G is a modified version of the Trier Social Stress Test (TSST, Kirschbaum, Pirke, and Hellhammer, 1993) that was designed to induce acute psychosocial stress. The TSST-G procedure and our modifications are described in more detail in Appendix 3.C.4.

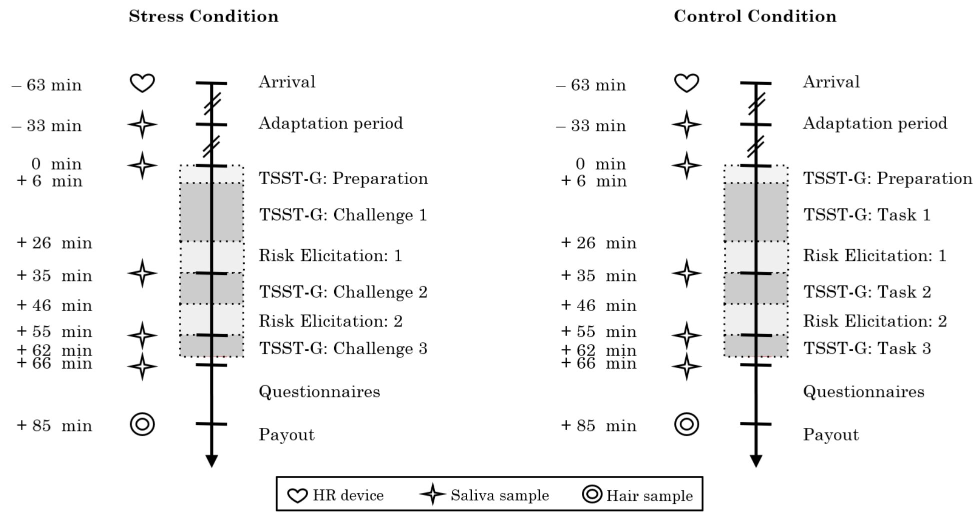


Figure 3.1. Timeline

To ensure credible estimates of the causal effect of acute stress on risk-taking, our within-subject design satisfies the balanced panel, temporal stability, and causal transience assumptions fundamental to valid within-subject studies (List, 2025). A detailed discussion of how these assumptions are met is provided in Appendix 3.C.1.

The following subsections and Appendix 3.C provide further details on the individual elements displayed in Figure 3.1, including the induced exogenous variation in stress, the risk elicitation procedure, physiological and self-reported stress measures, and further procedural details.[9]

3.2.2 Exogenous variation in *Stress*

Treatment phase in the Stress condition. In the *Stress* condition, we induced stress using an adjusted protocol of the TSST-G, adapted to simultaneously induce acute stress in several experiment participants.[10] It triggers acute stress by exposing individuals to a socio-evaluative threat.[11] Participants were confronted with a social-evaluative threat in three challenges: a mock job interview (conducted before the lottery choices), a complex arithmetic challenge, and a complex letter challenge. The order of the latter two challenges was randomized across sessions, with each challenge conducted either halfway through the lottery decisions or after all lottery decisions were completed. Participants performed all challenges in front of a com-

9. Appendix 3.C.3 presents the full chronological sequence of the experiment. The translated participant instructions are provided in Appendix 3.F.

10. See Appendix 3.C.4 for details.

11. Various studies that used this stress protocol have successfully induced acute stress (e.g., Buckert et al., 2014; Cahlíková and Cingl, 2017).

mittee consisting of one female and one male member wearing laboratory coats, and were filmed by video cameras.

Treatment phase in the Control condition. We implemented a control protocol along the lines of Von Dawans, Kirschbaum, and Heinrichs (2011), to best match the nature of the components of the *Stress* condition but without inducing any stress. The three tasks in the *Control* condition comprised reading a text (conducted before the lottery decisions), counting from one to 100, and reciting the alphabet (conducted halfway through and after the lottery choices, with the order randomized across sessions). During the tasks, participants stood in a row (segregated by their cubicles' walls) and conducted all tasks aloud and jointly with two experimenters (one female, one male).[12] Each participant was allowed to follow at their own pace.

3.2.3 Risk-taking

We used multiple price lists to elicit participants' certainty equivalents for 28 lotteries. We based our multiple price list design on previous risk-elicitation literature (see also Bruhin, Fehr-Duda, and Epper, 2010). For each row of a given multiple price list, participants decided whether they preferred receiving a safe payment of a certain amount or playing a risky lottery. Half of the 28 multiple price lists had outcomes in the gain domain, i.e., participants chose between risky and certain gains, while the other half had outcomes in the loss domain, i.e., participants chose between risky and certain losses. The gains ranged from 10 to 150 experimental points, while the losses ranged from -10 to -150 experimental points. The probability of realizing risky gains and losses varied between 5% and 95%. In line with Bruhin, Fehr-Duda, and Epper (2010), the multiple price lists presented safe payments in descending order (i.e., highest safe option on top) in the gain domain and in ascending order in the loss domain. Figure 3.2 shows an example of a decision in the gain domain, and Table 3.C.1 in Appendix 3.C.7 provides an overview of all lotteries used.

The multiple price lists were presented in randomized order and divided into two blocks of 14 decisions each. The first block was shown after the first, and the second block after the second challenge/task.

Participants completed the same multiple price lists in the *Stress* and *Control* conditions, and thus we observe the choices of a given participant for a given multiple price list twice, once in the presence and once in the absence of stress. This allows us to isolate the causal effect of acute stress in the analysis while keeping other decision characteristics constant.

To sustain an identical experiment timeline between all sessions and conditions, each multiple price list screen was timed at a maximum of 40 seconds. Participants

12. These two experimenters were different from the committee members in the *Stress* condition. The committee members in the *Stress* condition did not have any other role in the experimental procedures.

Decision-making situation 11: Gain situation

	Option A: Lottery	Your decision				Option B: Safe gain
1	Gain **20.00** points with **50%** probability. Gain **10.00** points with **50%** probability.	A	○	○	B	20.00 points
2		A	○	○	B	19.50 points
3		A	○	○	B	19.00 points
4		A	○	○	B	18.50 points
5		A	○	○	B	18.00 points
6		A	○	○	B	17.50 points
7		A	○	○	B	17.00 points
8		A	○	○	B	16.50 points
9		A	○	○	B	16.00 points
10		A	○	○	B	15.50 points
11		A	○	○	B	15.00 points
12		A	○	○	B	14.50 points
13		A	○	○	B	14.00 points
14		A	○	○	B	13.50 points
15		A	○	○	B	13.00 points
16		A	○	○	B	12.50 points
17		A	○	○	B	12.00 points
18		A	○	○	B	11.50 points
19		A	○	○	B	11.00 points
20		A	○	○	B	10.50 points
21		A	○	○	B	10.00 points

Next

Figure 3.2. Multiple price list: Gain situation

were informed that the computer would make a decision for them if they did not submit their decision on time. Decisions that timed out were not considered in our main analysis (compare Table 3.1), but we present an examination of the reasons for timeouts and check the robustness of our main results in Appendix 3.D.8.

We implemented an auto-fill function to facilitate the completion of multiple price lists and to reduce inconsistent choices. This function automatically completed the participants' choices based on logical dominance. Specifically, once a participant made a choice in one row, the system inferred and auto-filled all other rows where the preference could be clearly determined (assuming monotonicity and transitivity in preferences).[13] The auto-fill enforced a single switching point by design, helping reduce noise from inconsistent or erratic switching patterns. However, we did not completely eradicate the possibility of participants making stochastically dominated choices. Specifically, participants could still make stochastically dominated choices in the top or bottom rows of a lottery, as these always contained a first-order stochastically dominated option.[14]

13. For instance, if a participant chose the lottery over a safe payment of 15 points, the system automatically selected the lottery for all lower safe payments (i.e., less than 15 points), since choosing the lottery over 15 points implies a preference for the lottery at even lower guaranteed amounts. Conversely, the system did not make selections for rows with safe payments higher than 15 points, where the participant's preference remained ambiguous.

14. For example, in the multiple price list shown in Figure 3.2, choosing the lottery in the top row is dominated by choosing the safe gain of 20 points. Participants should always prefer a safe gain of 20

Participants' decisions were incentivized: at the end of each session, 1 of the 28 multiple price lists was randomly chosen for each participant, and the choice made in a randomly chosen row was implemented. The respective payment was added to or subtracted from participants' payout.

For each observed MPL, we construct a measure of how risk-seeking a participant behaves with respect to this specific lottery. First, we observe the row in which participants switch from preferring the safe payment to preferring the lottery (or vice versa).[15] This switching row identifies the interval in which the individual's certainty equivalent (CE) for the lottery must lie: bounded below by the highest safe amount they reject in favor of the lottery and above by the lowest safe amount they accept over the lottery.[16] We estimate the CE as the arithmetic mean of these bounds. Second, to enable comparisons between lotteries with different outcome magnitudes, i.e., different expected values (EV), we standardize the risk premium by the absolute value of the EV and define the standardized risk premium as $\overline{RP} = (EV - CE)/|EV|$.[17] According to this measure, a decision is classified as risk-averse if $\overline{RP} > 0$ and as risk-seeking if $\overline{RP} < 0$. Table 3.2 presents the summary statistics for the average standardized risk premia used in our main analysis.
Appendix 3.C.7 provides further information on our risk elicitation.

3.2.4 Stress measurement

To investigate whether we successfully manipulated the acute stress levels of participants, we measured acute stress in several ways throughout the experiment. We also measured participants' chronic stress levels.

Salivary cortisol. The hormone cortisol plays a key role in the stress response. It is released when the hypothalamic-pituitary-adrenal HPA) axis is activated in response to acute stress (Dedovic et al., 2009). Accordingly, cortisol levels are widely used as a biomarker of stress in experimental studies (Foley and Kirschbaum, 2010).[18] Since the release of free cortisol into the blood, i.e., biologically active cortisol, maps into

points to participating in a lottery that pays 20 points with 50% probability and 10 points with 50% probability. Similarly, choosing the safe gain of 10 points is dominated by choosing the lottery in the bottom row of the list.

15. Due to the use of an auto-fill mechanism in the choice interface, no participant displayed multiple switching points.

16. For illustration, consider the example shown in Figure 3.2. If a participant chooses the safe gain in rows 1-8 and the lottery in rows 9-21, the certainty equivalent must lie between 16 and 16.5 points.

17. Using the absolute values of the EV in the denominator allows for consistent interpretation across both gain and loss domains.

18. If a situation is perceived as stressful, it jump-starts a physiological stress response that involves activation of the HPA-axis, which initiates a cascade of reactions that result in the release of hormones, including cortisol. Cortisol induces physiological changes alerting the body and enabling it to bundle resources to deal with the threat (e.g., fight or flight).

a person's saliva (Kirschbaum and Hellhammer, 1994), cortisol levels can be conveniently tracked with saliva samples. Our saliva sampling procedures required participants to chew on a cotton swab (Salivette®, Sarstedt, Nuembrecht, Germany) for two minutes. Appendix 3.C.2 provides a detailed description of saliva sampling, explains the role of cortisol as a stress biomarker, and details how our research design addresses the challenges inherent in measuring it.

To track changes in salivary cortisol, we took a total of five saliva samples throughout the experiment.[19] Cortisol peaks in saliva between 20 and 30 minutes after the onset of a stressor and then falls subsequently until homeostasis is reached. Our first saliva sample tracks a potential stress response upon entering the laboratory. The second sample, taken before treatment onset, serves as a baseline measure of cortisol levels in the absence of stress, while samples three to five allow us to measure participants' responses to the treatment conditions and their cortisol levels during the risk elicitation.

Heart rate. We measured the heart rate in beats per minute during the experiment to account for stress and arousal due to the situation (Hjortskov et al., 2004). Von Dawans, Kirschbaum, and Heinrichs (2011), for example, found a significant heart rate response in participants who were treated by the stress protocol of the TSST-G. To measure heart rate, we equipped participants with a pulse sensor (Polar® Verity Sense), which is worn on the forearm or upper arm during the entire experiment.

Self-rated stress response. Finally, after the treatment phase, we surveyed subjective stress perceptions about the completed challenges by asking participants to rate their agreement with the statement 'How stressed did you feel?' on a Likert scale (see Dohmen, Rohde, and Stolp, 2023).

Measures of chronic stress. We also measured participants' chronic stress levels to assess whether and how the impact of acute stress differs by chronic stress level. To assess chronic stress, we use hair cortisol concentration (HCC), which has been shown to measure long-term exposure to stress (Stalder and Kirschbaum, 2012).[20]

At the end of their second session, participants could voluntarily provide a hair sample for a payment of 6 EUR. A hair sample consists of a strand of hair at least 2 cm long, cut near the scalp from the posterior vertex, close to the subcranial bone. This 2 cm hair segment contains the history of cortisol exposure during the past two months, based on an average hair growth rate of 1 cm/month (Wennig, 2000).

19. Specifically, we took samples 30, 60, 100, 120, and 131 minutes after the start of the experiment.

20. HCC does not measure acute stress but rather provides information on long-run exposure to stress. Importantly, hair cortisol has been found to reliably predict a wide range of psychological diseases such as major depression, post-traumatic stress, or generalized anxiety disorders, as well as acute myocardial infarction (see the reviews of Stalder and Kirschbaum, 2012; Staufenbiel et al., 2013, and the references therein) and hence provides us with an established biomarker to measure chronic stress.

HCC was analyzed for each of the two 1 cm sub-segments.[21] Seventy-one out of 156 participants provided us with hair samples, allowing us to measure their respective HCCs and providing a retrospective measure of participants' cortisol secretion over the last two months (Stalder and Kirschbaum, 2012).

As an additional measure of chronic stress, we surveyed subjective chronic stress using the Trier Inventory for the Assessment of Chronic Stress (TICS, Schulz and Schlotz, 1999), which aims to capture chronic stress throughout the last year. The TICS questionnaire was administered in participants' first session.

3.2.5 Procedural details

Our study was approved by the Ethics Committee of the Medical Faculty of the University of Bonn. Prior to participation, all individuals provided written informed consent. Participants were explicitly informed that their saliva samples would not be used for genetic or drug testing and received details on how their personal data, including heart rate, saliva, and video recordings, would be handled.

The experiment was conducted at the University of Bonn's Laboratory for Experimental Economics (BonnEconLab) between June and November 2022, across two waves. In total, 46 sessions were conducted, with each participant attending one *Stress* and one *Control* session. Sessions included up to 8 participants and were single-gender to control for possible gender-related effects. Each day included two sessions, starting at 1:30pm and 5:00pm, respectively.[22]

Recruitment was conducted via the hroot system (Bock, Baetge, and Nicklisch, 2014). Participants were informed in advance that saliva samples would be collected and required to comply with specific health and behavioral criteria (see Appendix 3.C.8). These precautions ensured the validity of cortisol-based stress measurements, as factors such as cortisone medication or intense physical activity can distort hormonal responses.

Each participant attended both sessions in the same group, with the sessions scheduled one week apart. We counterbalanced sessions so that roughly half of participants would experience their sessions in the early vs. late afternoon and roughly half would experience the *Stress* or *Control* condition first.[23]

The experiment was programmed in oTree (Chen, Schonger, and Wickens, 2016) and conducted in German. At the start of each session, oral instructions were provided, including a demonstration on how to use Salivettes® (Sarstedt, Nuembrecht, Germany) for saliva collection. Participants then proceeded to private cubicles to complete the computerized tasks.

21. Appendix 3.C.2 provides more details.

22. To control for circadian fluctuations in cortisol, participants completed both sessions either in the early (1:30pm) or late (5:00pm) time slot (see Kirschbaum and Hellhammer, 1989).

23. This design prevented systematic confounding between condition and session timing or order.

Across both sessions, participants spent approximately five hours in the laboratory and earned an average of EUR 70.37.[24] Earnings were accumulated as experimental points and converted to Euros at a rate of 15:1. Due to potential losses, all participants were endowed with 535 experimental points at the beginning of each session. Participants who opted to provide a hair sample received an additional EUR 6. All payments were made privately at the end of the second session.

3.2.6 Participants and data

A total of 156 individuals (mean age = 23.69 years; 50.64% female and 49.36% male, with no participants identifying as diverse; 97.44% students) completed both sessions. Following our preregistration, data from 142 participants (mean age = 23.73; 47.89% female; 98.59% students) were included in the main analysis.[25] Furthermore, our main analysis excludes decisions that were either not completed within the time limit or were first-order stochastically dominated. Our reasoning for the second exclusion restriction is that individuals were likely making errors in these choices (compare Dohmen and Gerasimou, 2024).[26] Additionally, our main analysis compares each individual's decision facing the *same* lottery in the *Stress* and *Control* conditions. We hence restrict our main dataset to instances in which we observe each participant's decision for the same lottery in both the *Stress* and *Control* conditions.[27] Following these exclusions, we obtain a final dataset of 6,810 individual choices, corresponding to 3,405 matched choice pairs, 1,721 in the gain domain, and 1,684 in the loss domain. Table 3.1 provides an overview of the application of the exclusion criteria and the resulting main analysis sample. Table 3.2 provides an overview of the average number of lottery decisions made by each participant. Table 3.2 also displays the average standardized risk premia of the respective decisions.

3.3 Results

In this Section, we present the results of our analyses. After providing evidence on the effectiveness of our stress induction procedure in the *Stress* condition in Section 3.3.1, we proceed in Section 3.3.2 to examining the effect of acute stress on risk-taking. Section 3.3.2.1 provides descriptive evidence. We then use within-subject analyses to examine whether risk-taking significantly differs between the *Stress* and

24. The average for the main analysis sample was EUR 70.33.

25. Our main results remain robust when we include individuals who were excluded from the main analysis because of a violation of behavioral rules or self-reported health conditions, which could be associated with the cortisol response, see Table 3.D.10 in Appendix 3.D.9.

26. Our main results remain robust when including these choices, see Table 3.D.11 in Appendix 3.D.9.

27. This restriction enhances both decision-level and individual-level analyses by ensuring that observed differences in risk-taking behavior are not confounded by differing sets of lottery choices across conditions.

Table 3.1. Application of exclusion criteria

Statistic	Value
Participant-level exclusions	
Total	156
Excluded due to behavioral rule violations	3
Excluded due to health reasons	10
Excluded due to comprehension check failure	1
Final sample	142
Decision-level exclusions	
Total after application of participant-level exclusions	7,952
- in stress condition	3,976
- in control condition	3,976
Number of dominated lottery decisions	197
- in stress condition	91
- in control condition	106
Number of timed-out lottery decisions	473
- in stress condition	222
- in control condition	251
Number of decisions without corresponding decision in other condition	472
- in stress condition	214
- in control condition	258
Final sample	6,810

Table 3.2. Summary statistics on final sample

Statistic	Value
Participant-level decision statistics	
Avg. number of lottery decisions taken	51.28 (7.57)
- in loss domain	25.48 (3.83)
- in gain domain	25.80 (4.00)
- in stress condition	25.80 (4.29)
- in control condition	25.49 (4.48)
Average standardized risk premium	0.014 (0.243)
- in loss domain	0.055 (0.245)
- in gain domain	-0.027 (0.234)
- in stress condition	0.014 (0.247)
- in control condition	0.014 (0.239)

Control conditions. We first test for a homogeneous shift toward more risk-seeking or risk-averse behavior (Section 3.3.2.2), and subsequently, whether risk-taking behavior shifts toward or away from risk neutrality (Section 3.3.2.3).

In our within-subject analyses, we analyze the impact of stress on risk-taking both at the lottery and at the participant level. In our lottery-level analyses, our unit of observation is at the participant-lottery level. Here, we compare the participants' standardized risk premia demanded in a given lottery between the *Stress* and *Control* conditions. This approach is deliberately theory-agnostic: it does not rely on any specific model of decision-making under risk, and it allows individuals to have varying degrees of risk aversion when facing different lotteries, making no assumption about consistency across choices.

In the participant-level analysis, we aggregate our lottery-level standardized risk premia to obtain a single measure of risk-taking for each participant. We assume that each participant can be characterized by a single, stable risk aversion parameter, for instance, as in expected utility theory.[28]

In contrast to several previous studies employing between-subject designs, we do not find evidence of a significant effect of acute stress on risk-taking. To understand whether this discrepancy owes to differences in our sample's composition or to the fact that our within-subject analysis approach better captures heterogeneities in baseline risk-taking that between-subject designs overlook, we additionally perform a between-subject analysis, presented in Section 3.3.2.4.

Finally, in Section 3.3.2.5, we provide an alternative approach to analyzing the heterogeneous effects of stress on risk-taking by conditioning effects on baseline risk-taking behavior in the *Control* condition. This approach yields differing results from the analysis in Section 3.3.2.3, which we trace back to methodological flaws in conditioning the analysis on baseline risk-taking. We provide these results, nevertheless, to illustrate the importance of a sound econometric approach and to caution future research.[29]

3.3.1 Assessing the stress induction

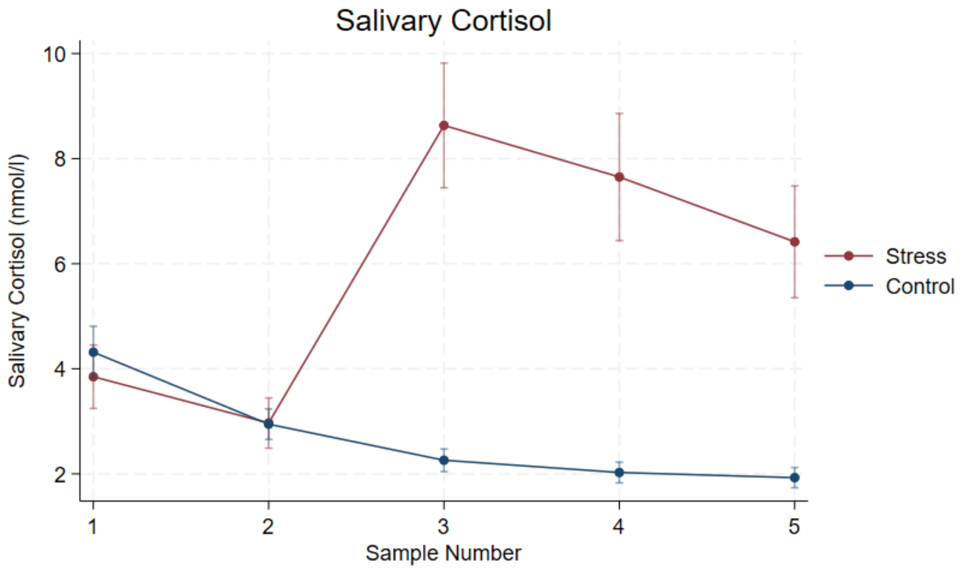
We begin by verifying that our *Stress* treatment induced a neuroendocrine stress response. Across all three physiological and subjective measures of acute stress, we find strong evidence that the stress induction procedure was effective.

Figure 3.3 demonstrates that the stress induction procedure significantly elevated cortisol levels, our primary indicator of stress. The cortisol trajectories reveal a marked divergence between the *Control* and *Stress* conditions. In the *Control* sessions, salivary cortisol concentrations follow the expected diurnal pattern (Kirschbaum and Hellhammer, 1989), decreasing steadily over the course of the experiment (Smyth et al., 1997).[30] In contrast, salivary cortisol of participants in the *Stress* condition rises markedly after the onset of the treatment phase, as is evident from a comparison of saliva Samples 2 and 3. Saliva Sample 2 was taken immedi-

28. In our estimations, we average over choices of an individual, assuming that individuals would make consistent choices in the absence of decision errors. Strictly speaking, a framework like expected utility theory (EUT) assumes that individuals make consistent choices under risk and fully comply with its core axioms. In practice, however, participants may make decision errors. Observed inconsistencies across all 28 lottery tasks could still be reconciled with the EUT framework, once such errors are taken into account. Averaging across all choices allows us to smooth out these errors. Hence, we do not assume perfect choice consistency.

29. We provide the results of further preregistered analyses in Appendix 3.D.9.

30. See Appendix 3.C.2 for more details on the diurnal cycle.



Notes: This figure depicts the average salivary cortisol, measured in nmol/l, in the *Stress* (red) and *Control* condition (blue). Samples 1 and 2 were collected before the onset of the TSST-G procedure. Samples 3, 4, and 5 were collected after the onset of the TSST-G procedure. The bars indicate the 95% confidence interval around the mean.

Figure 3.3. Salivary cortisol during the experiment

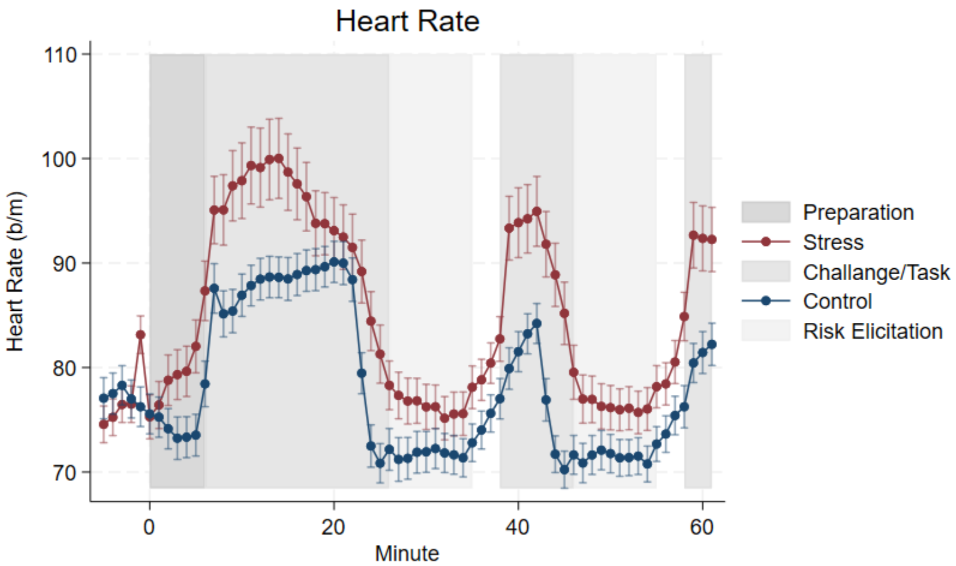
ately prior to the treatment phase, while Sample 3 was collected after the first challenge/task. Evidently, cortisol levels rise sharply: cortisol concentrations in Sample 3 are significantly elevated in the *Stress* condition relative to the *Control* condition (two-sided t-test, p-value < 0.001). On average, cortisol increased by 5.67 nmol/l in the *Stress* condition, while it declined by 0.69 nmol/l in the *Control* condition, following the diurnal cortisol pattern.[31] This significant difference persists across subsequent measurements. Sample 4, taken after the second challenge/task, and Sample 5, taken after the third challenge/task, both show significantly elevated cortisol levels in the *Stress* condition relative to the *Control* (both two-sided t-tests, p-value < 0.001).

As an additional manipulation check, we analyze participants' heart rate, another biological marker of acute stress.[32] Figure 3.4 illustrates heart rate patterns

31. Beyond aggregate effects, we assess individual stress responses during the TSST-G using a 2.5 nmol/l salivary cortisol increase threshold (Cahlíková and Cingl, 2017). In our sample, 59.86% of participants in the *Stress* condition are above this threshold, compared to none in the *Control* conditions, comparable to Cahlíková and Cingl (2017) and Parslow and Rose (2022). Kirschbaum, Pirke, and Hellhammer (1993) report a 70% rate among men in an individual TSST. Disaggregated by gender, 74% of males and 44% of females in the *Stress* condition meet the 2.5 nmol/l criterion, consistent with findings that women, while experiencing stress via other markers, often release less free cortisol (Uhart et al., 2006; Allen et al., 2014).

32. The heart rate analyses pertain to 282 observations, since there is no recording for two observations, due to a technical issue with one of the heart rate devices.

by treatment status throughout the treatment phase of the experiment. At the start of the preparation phase, indicated by minute 0 in Figure 3.4, the average heart rate does not differ between the *Stress* and the *Control* condition (two-sided t-test, p-value = 0.849).[33] However, following the onset of the treatment phase, heart rates in the *Stress* condition had increased significantly above those of the *Control* condition (two-sided t-test, p-value< 0.001). By the time participants had reached the middle of task one, i.e., 16 minutes into the treatment phase, the average heart rate had increased by 22.29 beats/min in the *Stress* condition, and 13.34 beats/min in the *Control* condition. This elevated heart rate persists throughout all challenges/tasks, risk elicitation blocks, and until the end of the session. Notably, these patterns are consistent between genders, with no significant gender differences in heart rate dynamics observed.[34]



Notes: This figure depicts the average heart rate in beats per minute during the experiment. It plots the heart rate measurements averaged for all participants in the *Stress* (red) and *Control* condition (blue). The bars indicate the 95% confidence interval around the mean.

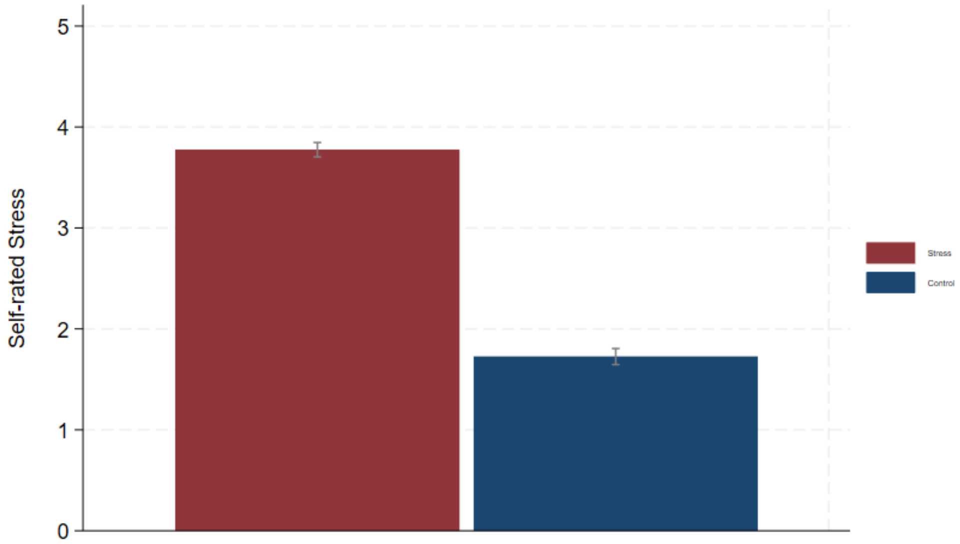
Figure 3.4. Heart rate during the experiment

Lastly, participants self-reported their perceived level of stress measured on a 5-point Likert scale, with results displayed in Figure 3.5 by treatment condition. Consistent with our physiological measures, perceived stress was significantly higher in the *Stress* condition than in the *Control* condition (mean_{Stress} = 3.77, $\text{mean}_{Control}$ =

33. In the immediate minutes before treatment onset there is a visible spike in the *Stress* condition, reflecting that these participants had to stand up and switch rooms, i.e., walk, for the task preparation.

34. Note that at the end of challenge one, the average heart rate falls more quickly in *Control* than in *Stress* because participants in the *Control* condition are allowed to sit down earlier than in the other condition for procedural reasons.

1.73; two-sided t-test, p-value < 0.001). We find no significant gender differences in self-reported stress within either condition (two-sided t-test: p-value$= 0.796$ and p-value$= 0.350$ in *Stress* and *Control*, respectively).



Notes: This figure depicts the averaged self-rated stress for all participants in the *Stress* (red) and *Control* condition (blue). The bars indicate the 95% confidence interval around the mean.

Figure 3.5. Self-rated stress

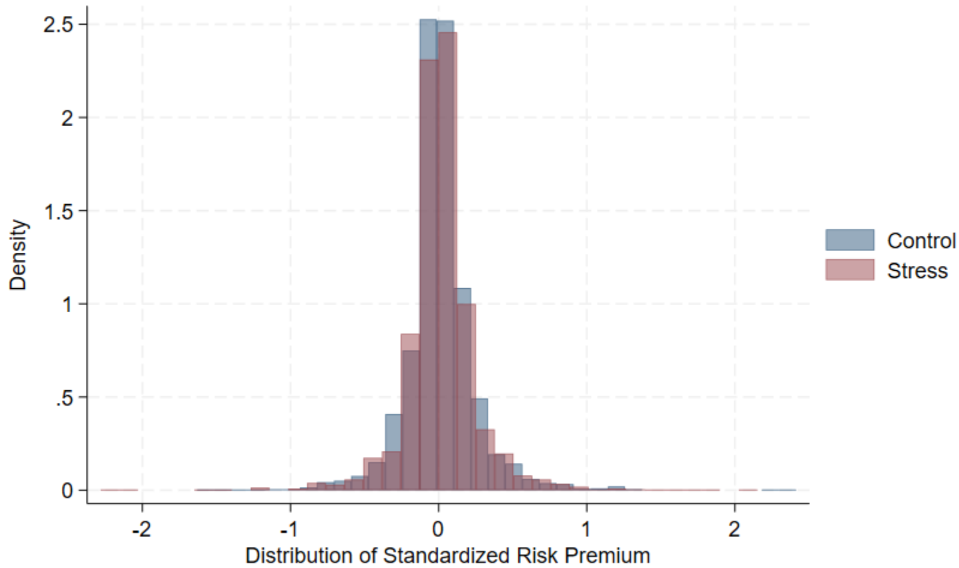
Overall, all three measures of acute stress confirm that the *Stress* treatment condition effectively induced a neuroendocrine stress response. Appendix 3.D.1 shows that there are no substantial differences with respect to our manipulation markers between participants who had the *Stress* treatment in their first session and those who had it in their second session.

3.3.2 Main analyses

Having established that the *Stress* treatment successfully induced a neuroendocrine stress response, we now turn to our main analysis, which examines the causal impact of stress on risk-taking behavior. As detailed in Section 3.2.3, we standardize the risk premium by the absolute value of the EV and define the standardized risk premium as $\overline{RP} = (EV - CE)/|EV|$. For the sake of brevity, we refer to the 'risk premium' in the following.

3.3.2.1 Descriptive analysis

Figure 3.6 plots the risk premia demanded in the *Stress* and *Control* conditions. Visually, there is no difference between these two conditions. A series of tests supports this insight: applying the conventional p-value of 0.05, we do not find a statistically significant difference between the means (two-sided Student's t-test, p-value = 0.916), nor the standard deviations (two-sample F-test for equality of variances, p-value = 0.057), nor in the distribution functions (two-sample KS-test, p-value = 0.822).



Notes: This figure depicts the standardized risk premium for all participants in the *Stress* (red) and *Control* condition (blue).

Figure 3.6. Distribution of the standardized risk premium

3.3.2.2 Within-subject analyses

Next, we leverage the advantages of our within-subject design to investigate how risk-taking behavior changes when participants are exogenously exposed to stress. We first analyze potential changes at the lottery level.

The outcome variable is the within-subject *change* in the risk premium demanded for a given lottery, i.e., the difference between participants' demanded risk premium in the presence of stress and the risk premium demanded in the absence of stress. This differencing approach controls by design for individual- and lottery-specific time-invariant factors that remain constant across both experimental sessions. These factors are, for instance, individuals' underlying risk preferences as well as the domain, skewness, and expected payoff of the lottery. A positive value of $\Delta\overline{RP}_{il}$ indicates a shift toward greater risk aversion under stress, while a negative value reflects increased risk-taking.

We estimate the following equation:

$$\Delta\overline{RP}_{il} = \alpha + \varepsilon_{il}, \tag{3.1}$$

where $\Delta\overline{RP}_{il}$ is the *change* in risk premium demanded by individual i for lottery l, between the *Stress*, $\overline{RP}_{il,Stress}$, and *Control*, $\overline{RP}_{il,Control}$, conditions. The constant term α captures whether this term is, on average, positive or negative, capturing any systematic move toward risk aversion or risk seeking. Participants attended both of their sessions within the same group of participants; hence, the order of the *Control* and *Stress* condition sessions was assigned at the group level. We thus cluster standard errors at the group level as the level of treatment assignment (Abadie et al., 2023).[35] Hence, our main analysis is equivalent to a t-test clustering standard errors at the group level. Our differencing approach yields the same results as directly regressing the demanded risk premium on a *Stress* condition dummy, including individual and lottery fixed effects and clustering standard errors (compare Table 3.D.2 in Appendix 3.D.3).[36]

The estimation results are presented in Table 3.3. Column (1) estimates Equation (3.1), clustering standard errors at the group level. We do not find evidence of a significant effect of stress on the demanded risk premium. Notably, the coefficient we estimate is quite small: ~ 0.001 and 4% of the demanded risk premium averaged across individuals and decisions.[37] Columns (2) and (3) restrict the analysis to the gain and the loss domains,[38] respectively, and find similar results. While the analyses in Columns (1)-(3) are at the lottery level, Column (4) displays the results of analyses at the participant level.[39] To this end, we average $\overline{RP}_{il,Stress}$ and $\overline{RP}_{il,Control}$ for each participant, and then test for a change between these two averages by defining an outcome variable, which is their difference. Here, we also do not find any significant effects, with an estimated coefficient of 0.002.

Importantly, we find no evidence that acute stress affects risk-taking, and this result is robust to a variety of sensitivity checks. First, our results do not differ depending on whether participants experienced the *Stress* condition in their first or second session (see Table 3.D.3 in Appendix 3.D.4).

Next, instead of focusing on the standardized risk premium, we assess whether stress induces systematic changes in the MPL row, in which participants' preferences change from lottery to safe payment or vice versa, i.e., their 'switching row'. Directly analyzing the switching row instead of the certainty equivalent (CE) allows for a more straightforward analysis, since it does not require any assumption on participants' true CE, which we assume in the main analysis to lie in the middle of an interval bounded by the safe payoffs of the switching row and the preceding row.

35. Table 3.D.1 in Appendix 3.D.2 shows that alternative clustering approaches (individual level, lottery level, no clustering) yield qualitatively similar results.

36. While we preregistered the analysis shown in Appendix 3.D.3, we prefer to display here the analysis according to Equation (3.1) for ease of exposition.

37. Back of the envelope calculation: $0.0006/0.014 = 0.042$. The Cohen's D of the effect is ≈ 0.003

38. We preregistered a separate analysis by loss and gain domain.

39. We preregistered a separate analysis at both the lottery and the participant level.

Table 3.3. Within-subject comparison in risk-taking

	Change in Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Constant	0.001	-0.002	0.003	0.002
	(0.005)	(0.006)	(0.006)	(0.005)
Observations	3405	1721	1684	142
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.000	0.000	0.000	0.001

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The analysis is at the participant-decision level in Columns (1) - (3), and at the participant-level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

Figure 3.D.1 in Appendix 3.D.5 plots the difference between the switching row in *Stress* and *Control* per lottery and participant over all observations used for the main analysis. We find that the same switching row occurs for 26% of all lottery pairs, and that the majority of lottery pairs (55%) show a difference within one row, which is consistent with the absence of a systematic stress effect on risk-taking.

Further, Appendix 3.D.6 estimates different parametric models of risk-taking behavior, derived from expected utility theory, assuming a power utility function (as in Bruhin, Fehr-Duda, and Epper, 2010; Murad, Sefton, and Starmer, 2016), as well as prospect theory, using a linear in log odds function (as in Goldstein and Einhorn, 1987; Lattimore, Baker, and Witte, 1992). Estimates do not substantially differ depending on whether we use data from the *Stress* or *Control* sessions to estimate our models, supporting the rejection of a substantial change between session types.[40]

Finally, previous research has shown that chronic stress, defined as stress sustained over extended periods, influences how individuals respond to acute stressors (Chida and Hamer, 2008). Hence, our main results could deviate for chronically stressed individuals. We therefore provide a detailed analysis using multiple physiological and subjective measures of chronic stress in Appendix 3.D.7. It is important to note that our sample does not comprise individuals with pathological levels of chronic stress. Within this non-pathological range, we observe almost no heterogene-

40. Such an analysis sets out to infer a risk preference parameter from task performance, assuming a specific functional form for risk preferences. To this end, we estimate and compare the parameters that best fit the observed risk-taking behavior in the *Stress* and *Control* conditions, respectively. This approach, of course, hinges on the assumption that the chosen functional form reasonably approximates individuals' true preferences.

ity in the null effect of chronic stress,[41] but this result may not generalize to more extreme cases.

3.3.2.3 Is the effect of stress contingent on baseline risk-taking?

While we do not find evidence of increased risk aversion or increased risk seeking under stress, the effect of acute stress may also be more nuanced. We preregistered the hypothesis that the effect of stress on risk-taking is contingent on an individual's risk-taking behavior when not stressed. Two possibilities emerge: First, acute stress might make previously risk-averse decisions more risk-seeking, while making previously risk-seeking decisions more risk-averse. This implies that acute stress shifts decisions toward risk neutrality. A second possibility is that acute stress intensifies baseline tendencies in risk-taking behavior, i.e., making risk-averse behavior more risk-averse and risk-seeking behavior more risk-seeking. In both cases, such patterns would not or only marginally (depending on baseline composition) affect the average risk premium demanded, and are thus not detectable in a regression analysis based on Equation (3.1).

We thus estimate a second within-subject regression:

$$\Delta|\overline{RP}_{il}| \;=\; \alpha + \varepsilon_{il}. \tag{3.2}$$

The dependent variable in Equation (3.2) is the change in the *absolute* value of the risk premium, i.e., calculated as the difference between the *absolute* risk premium in *Stress* and the *absolute* risk premium in *Control*. If acute stress were to influence risk-taking behavior by shifting decisions toward risk-neutrality (the first pattern described in the paragraph above), this would be visible in a decrease in the absolute risk premium demanded. If acute stress were to shift decisions away from risk-neutrality (the second pattern), this would be visible in an increase in the absolute risk premium demanded. Again, we do not find any evidence for either effect. The coefficient in Column (1) is estimated at 0.002, 1% of the average absolute risk premium demanded (0.15).[42] Columns (2) and (3) find similar results examining the gain and loss domains separately.[43]

Column (4) again switches gears, examining the relationship in a participant-level analysis. We average the absolute risk premium demanded across each individual,[44] estimating an insignificant coefficient of 0.001.

41. An exception is one of the six sub-scales of our subjective chronic stress measure, which seems to suggest some heterogeneity in treatment effects depending on self-reported experience of social strain within the past year. However, given that the other five sub-scales do not show evidence of a heterogeneous effect and that there is no other supporting evidence in the data, we refrain from further interpreting this pattern.

42. The Cohen's D of the effect is ≈ 0.009

43. We preregistered a separate analysis by loss and gain domain.

44. We preregistered a separate analysis on both the lottery and on the participant level.

Table 3.4. Within-subject comparison in absolute measure of risk-taking

	Change in Absolute Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Constant	0.002	0.007	-0.004	0.001
	(0.005)	(0.005)	(0.007)	(0.006)
Observations	3405	1721	1684	142
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.000	0.002	0.000	0.000

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the absolute standardized risk premium (RP, see Section 3.2.3), i.e., the difference between the absolute risk premia demanded in the *Stress* and *Control* conditions. The analysis is at the participant-decision level in Columns (1) - (3), and at the participant level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.3.2.4 Between-subject analysis

To compare our results to previous literature that has often used a between-subject analysis, we also offer a between-subject analysis below. To this end, we only use data from participants' first session as this resembles a clean between-subject design, drawing on a sample size comparable to that of previous studies.[45] Specifically, we compare the risk-taking behavior of participants who were assigned to the *Stress* condition in their first session with the risk-taking behavior of participants who were assigned to the *Control* condition in their first session. Thereby, we can provide a conventional between-subject estimate of the average treatment effect of stress on risk-taking behavior for our sample of participants.

We estimate the following equation:

$$\overline{RP}_{isl} = \alpha + \beta_1 Stress_{is} + X_{il} + \varepsilon_{isl}, \tag{3.3}$$

where $\overline{RP}_{isl}$ is the risk premium demanded by individual i in session s for lottery l. $Stress_{is}$ is an indicator whether individual i was in the *Stress* condition in session s, and X_{il} are controls at the individual- and lottery-level. Again, we cluster standard errors at the group level.
The results of this analysis are displayed in Table 3.5.

The coefficient estimate for β_1 in Column (1) is -0.029 (p-value $= 0.032$), indicating that individuals take, on average, less risk-averse decisions under stress. Adding controls for participants' age and gender (Column (2)) barely changes the

45. See also our preregistered analysis.

Table 3.5. Between-subject comparison in risk-taking (first-week sessions)

	Standardized RP				
	(1) (Full)	(2) (Full)	(3) (Full)	(4) (Gain)	(5) (Loss)
Stress	-0.029**	-0.028**	-0.031**	-0.037**	-0.025
	(0.013)	(0.012)	(0.013)	(0.014)	(0.016)
Female		0.014	0.017	0.007	0.027
		(0.012)	(0.013)	(0.014)	(0.016)
Age		0.000	0.000	-0.001	0.002
		(0.002)	(0.001)	(0.001)	(0.002)
Constant	0.030***	0.019	-0.142***	-0.097**	-0.138**
	(0.007)	(0.037)	(0.036)	(0.035)	(0.054)
Lottery FE			X	X	X
Observations	3405	3405	3405	1721	1684
R^2	0.004	0.004	0.448	0.442	0.426

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the standardized risk premium (RP, see Section 3.2.3). Columns (1) - (3) report estimates for the full sample, Column (4) reports estimates for the gain domain, and Column (5) for the loss domain. The specifications in Columns (2) - (5) control for a female dummy and participants' age, and Columns (3) - (5) additionally include lottery fixed effects. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

estimated treatment effect to -0.028 (p-value = 0.033). Next, in Column (3), we add lottery fixed effects, which account for lottery-specific characteristics. While the main coefficient of interest remains qualitatively unaffected, the explanatory power of the model largely increases, implying that specific lottery attributes affect the RP that participants demand.

While the between-subject analysis suggests a difference in risk-taking between stressed and non-stressed participants, the absence of a corresponding effect in the better-powered within-subject analysis indicates that this difference is unlikely to be caused by acute stress itself and is instead consistent with preexisting differences in baseline risk preferences between treatment and control groups.[46] Such differences may also be one reason for observing contradictory results in previous between-subject studies. Our within-subject analysis does control for these differences by design. Importantly, the presence of a between-subject difference in our

46. As an imperfect test of this, we compare the average risk premia chosen in *Control* between participants, who experienced the *Control* condition first, and those who experienced it as their second session, i.e., the individuals who are in the *Stress* condition in our between-subject analysis. The difference between these two groups is 0.033, i.e., roughly in line with the coefficient estimate for the effect of stress on risk-taking in our between-subject analysis. This difference is also statistically significant (two-sided t-test, p-value < 0.01), thereby supporting the notion that there are baseline differences between these two groups.

own data suggests that our null within-subject result is not driven by an unusually selected or unresponsive sample relative to earlier studies. Rather, the divergence appears to reflect the additional control and statistical power offered by the within-subject design, which eliminates baseline heterogeneity by construction. By increasing statistical power, this design strengthens the credibility of our null result. Table 3.B.1 in Appendix 3.B illustrates this advantage by comparing our experiment's power to detect effect sizes reported in previous studies under both between-subject and within-subject specifications.

Relying on a sample size common in the literature, our between-subject analysis illustrates how observed differences in risk-taking behavior between the *Stress* and *Control* groups can be mistakenly attributed to a treatment, even when they likely reflect a combination of differences in latent baseline risk preferences across individuals and limited randomization in small samples.

3.3.2.5 Alternative analysis conditioning on baseline risk-taking

An intuitive and straightforward alternative to the approach in Equation (3.2) to test whether the effect of stress on risk-taking depends on an individual's baseline propensity to take risks in a given lottery is to estimate whether $\Delta\overline{RP}_{il}$ systematically differs depending on whether the respective lottery choice was risk-averse or risk-seeking in the *Control* session. To this end, one could define $RiskAverse_{il}$, which equals 1 if individual i made a risk-averse choice in lottery l in the *Control* condition, and $RiskSeeking_{il}$, which equals 1 if individual i made a risk-seeking choice in lottery l in the *Control* condition, and then estimate the change in risk-taking under stress according to the following equation:

$$\Delta\overline{RP}_{il} = \alpha + \beta_1 RiskAverse_{il} + \beta_2 RiskSeeking_{il} + \varepsilon_{il} \tag{3.4}$$

The coefficient estimates of an OLS regression of Equation (3.4), shown in Table 3.A.2 in Appendix 3.A, point to a heterogeneous effect of stress on risk-taking: for decisions in which an individual behaves risk-averse in the absence of stress, the demanded risk premium seems to decrease in the presence of stress. Likewise, for decisions in which an individual behaves risk-seeking in the absence of stress, the demanded risk premium is estimated to increase. This pattern suggests a shift toward risk neutrality.

However, these results are ill-founded because conditioning on observed risk-taking in the *Control* condition produces results that are likely compromised by measurement error and resulting regression to the mean. The intuition for this is as follows: even if we assume individuals to have a set of stable latent risk preferences, it is plausible that individuals make errors in their decisions, resulting in (sometimes positive and sometimes negative) measurement error.

If individuals' choices reveal an overstated risk premium in one elicitation because of a positive draw from a measurement error distribution with mean zero, their choices will be expected to reveal a lower risk premium in a second elicitation.[47] Correspondingly, if we classify individuals based on the preference revealed by their possibly overstated (understated) risk premium in the *Control* condition, we would observe on average a less overstated (understated) risk premium in the *Stress* condition, even if there was no effect of stress on risk-taking behavior. As a result, on average, those who are classified as risk-averse based on the noisy decision in *Control* would appear to become less risk-averse; and those who are classified as risk-seeking, on average, would appear to become less risk-seeking in the *Stress* condition.

To determine whether the observed result reflects a genuine shift toward risk-neutral behavior or is merely an artifact of the measurement error described above, we 'flip' the analysis and base the classification of the independent variable on risk-taking behavior in the *Stress* condition (see Table 3.A.3 in Appendix 3.A). If stress truly induced risk-neutrality, we would observe in the 'flipped analysis' that the absence of stress drives individuals away from risk-neutrality. However, in the 'flipped analysis', it appears that the absence of stress rather than the presence of stress drives people toward risk-neutral behavior. This contradiction provides evidence that the shift toward risk-neutrality we originally observed is an artifact of measurement error, consistent with the main results in Table 3.4.

3.4 Concluding discussion

Using a novel within-subject laboratory experiment, we investigate how stress influences individual risk-taking behavior. We neither find evidence that acute stress increases or decreases risk aversion, nor that it shifts behavior toward or away from risk neutrality. We demonstrate that when applying methods common in this strand of literature, we find that our data can produce different conclusions. This pattern illustrates that previously diverging findings in the literature may in part be attributable to small sample sizes and measurement error.

The absence of evidence that stress either increases or decreases risk-taking differs from findings in several previous studies that rely on between-subject designs. Strikingly, in a between-subject analysis similar to the previous studies, we also estimate a statistically significant effect of stress on risk-taking. Our better-powered within-subject analysis, however, finds no statistically significant effect.[48] A likely

47. Suppose an individual's expressed risk premium is subject to symmetric measurement error around the true risk premium (mean). If an observed risk premium is below the true premium (a negative error), the expected value of the next observation is higher than the current draw, reverting back to the mean.

48. Table 3.B.1 in Appendix 3.B compares the power of our within-subject analysis and between-subject analysis to detect effect sizes common in the literature.

explanation for this discrepancy lies in heterogeneity in participants' latent risk preferences: since risk preferences vary across individuals, a relatively small sample size (as common in the existing literature) may, by chance, create alleged differences between treatment groups.[49] Our within-subject analysis, in contrast, by design controls for heterogeneity in individual risk preferences and thereby yields higher statistical power.

In addition, we do not find evidence of stress shifting risk-taking toward or away from risk neutrality, adding a new perspective on our understanding of risk-taking behavior. Specifically, given previous results and the neurological effects of stress, a plausible explanation for the previously mixed findings on the effect of stress on risk-taking could have been that the effect is heterogeneous: stress might push risk-averse decisions toward risk seeking and risk-seeking decisions toward risk aversion. We provide first evidence on a possible shift from or toward risk neutrality by constructing an outcome variable that directly captures a decision's distance from risk neutrality. We do not find any evidence for such patterns, making it unlikely that the inconclusiveness of the previous results was driven by heterogeneity in the relationship between acute stress and risk-taking, rather than the unfortunate combination of measurement error and sample size outlined above.

The presence of measurement error in capturing individuals' risk-taking attitudes implies that the methodology used for heterogeneity analyses must be chosen with extreme caution. Specifically, while conditioning on individuals' behavior as measured at baseline seems the most intuitive approach to assess potential heterogeneities driven by baseline behavior, our results illustrate that this approach produces spurious treatment effects. It is well known that measurement error in baseline variables leads to attenuation bias in regression estimates. However, researchers still often assess treatment effects across groups defined by their baseline behavior, even if these are measured with noise.

In the context of our study, the most intuitive approach to analyze heterogeneities by baseline behavior is to compare within-subject differences in risk-taking between the *Stress* and *Control* conditions across groups classified by their risk-taking behavior in the *Control* condition. We show that this approach is flawed because baseline risk-taking is measured with noise, such that an analysis yields misleading results. This methodological concern is not restricted to our setting but also arises in other contexts. For instance, estimates of the effectiveness of educational programs targeted at low-achieving students can be flawed if program participation is assigned based on students' noisily measured initial performance.[50]

49. Notably, two large-sample studies using a between-subject design, Parslow and Rose (2022) and Veszteg et al. (2021), both report null effects of acute stress on risk-taking.

50. Other examples include evaluating personnel policies across productivity groups when baseline productivity is measured with error, or assessing a regional growth policy if the policy measure is implemented in specific regions based on their noisily measured GDP.

From a methodological perspective, our study highlights three key considerations: (1) the importance of an adequate experimental design when examining the effects of stress on risk-taking–or, more generally, of visceral factors on behavior, preferences, traits, or beliefs; (2) the need to account for measurement error; and (3) the role of statistical power.

In light of our well-powered null results and the mixed findings in the existing literature, we conclude with considerable confidence that acute stress is unlikely to causally affect risk-taking in small-stakes financial decisions. To be concrete about how our results update beliefs about the existence of such an effect, consider the following post-study probability calculation. Based on the literature, a reasonable prior probability of a causal effect of acute stress on risk-taking is in the range of 25-75% (see Appendix 3.B for details). Given our null result, and noting that our study is powered to detect all significant effect sizes reported in previous studies with at least 80% power (see Table 3.B.1), we update a prior of 25% (the lower bound of our range) to a post-study probability of approximately 7%, and a prior of 75% (the upper bound of our range) to a post-study probability of approximately 39%. This substantial downward revision indicates that our null result meaningfully reduces the likelihood of a causal effect of acute stress on risk-taking (Maniadis, Tufano, and List, 2014; Abadie, 2020; List, 2026).

In terms of practical implications, our evidence boosts confidence that acute stress does not causally drive individual fluctuations in risk-taking behavior, especially in small-stakes financial decisions. This central result has practical relevance, for example, for research on delegation and decentralization (Aghion and Tirole, 1997), which largely does not consider the psychological and physiological state of the agent. Our study complements this literature by examining whether the acute psychological and physiological conditions of the decision maker affect choice behavior. Our results suggest that acute stress does not introduce additional variability that principals need to consider when deciding whether to delegate risky financial decisions. For organizations, this insight provides an important implication: supervisors can delegate many decisions without concern that acute stress, which is pervasive in many work environments, will systematically alter employees' decisions.

More generally, our study highlights that credible policy inference requires careful attention to measurement error, adequate sample size, and a study design that isolates treatment effects from underlying heterogeneity.

References

Abadie, Alberto. 2020. “Statistical nonsignificance in empirical economics.” *American Economic Review: Insights* 2 (2): 193–208. [143, 157]

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2023. “When should you adjust standard errors for clustering?” *Quarterly Journal of Economics* 138 (1): 1–35. [135]

Aghion, Philippe, and Jean Tirole. 1997. “Formal and real authority in organizations.” *Journal of Political Economy* 105 (1): 1–29. [143]

Allen, Andrew P., Paul J. Kennedy, John F. Cryan, Timothy G. Dinan, and Gerard Clarke. 2014. “Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test.” *Neuroscience & Biobehavioral Reviews* 38: 94–124. [131]

Bendahan, Samuel, Lorenz Goette, J. Thoresen, Leyla Loued-Khenissi, Fiona Hollis, and Carmen Sandi. 2017. “Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk.” *European Journal of Neuroscience* 45 (7): 877–85. [117, 118, 149, 153, 155, 157]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. “hroot: Hamburg registration and organization online tool.” *European Economic Review* 71: 117–20. [127]

Bozovic, Djordje, Maja Racic, and Nedeljka Ivkovic. 2013. “Salivary cortisol levels as a biological marker of stress reaction.” *Med Arch* 67 (5): 374–77. [159]

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper. 2010. “Risk and rationality: Uncovering heterogeneity in probability distortion.” *Econometrica* 78 (4): 1375–412. [121, 123, 136, 170, 176–178]

Buckert, Magdalena, Christiane Schwieren, Brigitte M. Kudielka, and Christian J. Fiebach. 2014. “Acute stress affects risk taking but not ambiguity aversion.” *Frontiers in Neuroscience* 8: 82. [117, 122, 149, 153, 155, 157]

Cahlíková, Jana, and Lubomír Cingl. 2017. “Risk preferences under acute stress.” *Experimental Economics* 20 (1): 209–36. [118, 122, 131, 149, 153, 155, 157, 186]

Ceccato, Smarandita, Brigitte M. Kudielka, and Christiane Schwieren. 2016. “Increased risk taking in relation to chronic stress in adults.” *Frontiers in Psychology* 6: 1–13. [178]

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. “oTree – An open-source platform for laboratory, online, and field experiments.” *Journal of Behavioral and Experimental Finance* 9: 88–97. [127]

Chida, Yoichi, and Mark Hamer. 2008. “Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: A quantitative review of 30 years of investigations.” *Psychological Bulletin* 134 (6): 829. [136, 178]

Cingl, Lubomír. 2019. “Stress and decision making in economic research.” Habilitation thesis. Faculty of Economics, University of Economics, Prague. [165]

Clark, Luke, Ruoran Li, Chelsea M. Wright, Farima Rome, Graeme Fairchild, Barnaby D. Dunn, and Michael R.F. Aitken. 2012. “Risk-avoidant decision making increased by threat of electric shock.” *Psychophysiology* 49 (10): 1436–43. [118, 149, 153, 156, 157]

Cohn, Alain, Jan Engelmann, Ernst Fehr, and Michel André Maréchal. 2015. “Evidence for countercyclical risk aversion: An experiment with financial professionals.” *American Economic Review* 105 (2): 860–85. [149, 153, 155, 157]

Correia, Sergio. 2014. “reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects.” Statistical Software Components S457874. Department of Economics, Boston College. [174]

Correia, Sergio. 2016. “reghdfe: Estimating linear models with multi-way fixed effects.” In *2016 Stata Conference.* 24. Stata Users Group. [174]

Costa, P. T., and R. R. McCrae. 1985. “The NEO personality inventory: Manual, form S and form R.” *Psychological Assessment Resources.* [171]

Cueva, Carlos, R. Edward Roberts, Tom Spencer, Nisha Rani, Michelle Tempest, Philippe N. Tobler, Joe Herbert, and Aldo Rustichini. 2015. “Cortisol and testosterone increase financial risk taking and may destabilize markets.” *Scientific Reports* 5 (1): 11206. [118, 149, 153, 155, 157]

Daughters, Stacey B., Stephanie M. Gorka, Alexis Matusiewicz, and Katelyn Anderson. 2013. “Gender specific effect of psychological stress and cortisol reactivity on adolescent risk taking.” *Journal of Abnormal Child Psychology* 41 (5): 749–58. [118, 149, 153, 155, 157]

Dedovic, Katarina, Miriam Rexroth, Elisabeth Wolff, Annie Duchesne, Carole Scherling, Thomas Beaudry, Sonja Damika Lue, Catherine Lord, Veronika Engert, and Jens C. Pruessner. 2009. “Neural correlates of processing stressful information: An event-related fMRI study.” *Brain Research* 1293: 49–60. [125]

Dettenborn, L., A. Tietze, C. Kirschbaum, and T. Stalder. 2012. “The assessment of cortisol in human hair: Associations with sociodemographic variables and potential confounders.” *Stress* 15 (6): 578–88. [179]

Dickerson, Sally S., and Margaret E. Kemeny. 2004. “Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research.” *Psychological Bulletin* 130 (3): 355. [160]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2008. “Representative trust and reciprocity: Prevalence and determinants.” *Economic Inquiry* 46 (1): 84–90. [171]

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. “Homo reciprocans: Survey evidence on behavioural outcomes.” *Economic Journal* 119 (536): 592–612. [171]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. “Individual risk attitudes: Measurement, determinants, and behavioral consequences.” *Journal of the European Economic Association* 9 (3): 522–50. [171]

Dohmen, Thomas, and Georgios Gerasimou. 2024. “Learning to maximize (expected) utility.” *ArXiv preprint arXiv:2402.16538.* [128]

Dohmen, Thomas, and Tomáš Jagelka. 2024. “Accounting for individual-specific reliability of self-assessed measures of economic preferences and personality traits.” *Journal of Political Economy Microeconomics* 2 (3): 399–462. [151]

Dohmen, Thomas, Ingrid Rohde, and Tom Stolp. 2023. “Tournament incentives affect perceived stress and hormonal stress responses.” *Experimental Economics* 26 (4): 955–85. [121, 126, 160, 161, 171]

Dreyer, Anna J., Dale Stephen, Robyn Human, Tarah L. Swanepoel, Leanne Adams, Aimee O’Neill, W. Jake Jacobs, and Kevin G. F. Thomas. 2022. “Risky decision making under stressful conditions: Men and women with smaller cortisol elevations make riskier social and economic decisions.” *Frontiers in Psychology* 13: 810031. [149, 153, 154, 157]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. “Global evidence on economic preferences.” *Quarterly Journal of Economics* 133 (4): 1645–92. [171]

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. “The preference survey module: A validated instrument for measuring risk, time, and social preferences.” *Management Science* 69 (4): 1935–50. [151]

Fehr, Ernst, and Klaus M. Schmidt. 2006. "The economics of fairness, reciprocity and altruism – Experimental evidence and new theories." *Handbook of the Economics of Giving, Altruism and Reciprocity,* 615–91. [171]

Finy, M. Sima, Konrad Bresin, Donna L. Korol, and Edelyn Verona. 2014. "Impulsivity, risk taking, and cortisol reactivity as a function of psychosocial stress and personality in adolescents." *Development and Psychopathology* 26 (4): 1093–111. [149, 153, 155, 157]

Foley, Paul, and Clemens Kirschbaum. 2010. "Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings." *Neuroscience & Biobehavioral Reviews* 35 (1): 91–96. [125]

Gao, Wei, Tobias Stalder, Paul Foley, Manfred Rauh, Huihua Deng, and Clemens Kirschbaum. 2013. "Quantitative analysis of steroid hormones in human hair using a column-switching LC–APCI–MS/MS assay." *Journal of Chromatography B* 928: 1–8. [162]

Gathmann, Bettina, Frank P. Schulte, Stefan Maderwald, Mirko Pawlikowski, Katrin Starcke, Lena C Schäfer, Tobias Schöler, Oliver T. Wolf, and Matthias Brand. 2014. "Stress and decision making: Neural correlates of the interaction between stress, executive functions, and decision making under risk." *Experimental Brain Research* 232 (3): 957–73. [149]

Gerlitz, Jean-Yves, and Jürgen Schupp. 2005. "Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP." Research Notes 4. DIW Berlin. [171]

Goldstein, Warren M., and Hillel J. Einhorn. 1987. "Expression theory and the preference reversal phenomenon." *Psychological Review* 94 (2): 236–54. [136, 176]

Hjortskov, Nis, Dag Rissén, Anne Katrine Blangsted, Nils Fallentin, Ulf Lundberg, and Karen Søgaard. 2004. "The effect of mental stress on heart rate variability and blood pressure during computer work." *European Journal of Applied Physiology* 92 (1-2): 84–89. [126]

Kandasamy, Narayanan, Ben Hardy, Lionel Page, Markus Schaffner, Johann Graggaber, Andrew S. Powlson, Paul C. Fletcher, Mark Gurnell, and John Coates. 2014. "Cortisol shifts financial risk preferences." *Proceedings of the National Academy of Sciences* 111 (9): 3608–13. [149, 178]

Kirschbaum, Clemens, and Dirk H. Hellhammer. 1989. "Salivary cortisol in psychobiological research: An overview." *Neuropsychobiology* 22 (3): 150–69. [127, 130]

Kirschbaum, Clemens, and Dirk H. Hellhammer. 1994. "Salivary cortisol in psychoneuroendocrine research: Recent developments and applications." *Psychoneuroendocrinology* 19 (4): 313–33. [126, 160]

Kirschbaum, Clemens, Karl-Martin Pirke, and Dirk H. Hellhammer. 1993. "The 'Trier Social Stress Test' – A tool for investigating psychobiological stress responses in a laboratory setting." *Neuropsychobiology* 28 (1-2): 76–81. [121, 131]

Kluen, Lisa Marieke, Agorastos Agorastos, Klaus Wiedemann, and Lars Schwabe. 2017. "Cortisol boosts risky decision-making behavior in men but not in women." *Psychoneuroendocrinology* 84: 181–89. [149, 153, 154, 157]

Kudielka, Brigitte M., Roland von Känel, Daniel Preckel, Lilian Zgraggen, Katharina Mischler, and Joachim E. Fischer. 2006. "Exhaustion is associated with reduced habituation of free cortisol responses to repeated acute psychosocial stress." *Biological Psychology* 72 (2): 147–53. [178]

Lattimore, Philip K., Robert R. Baker, and Gregory Witte. 1992. "The influence of probability on risky choice: A parametric examination." *Journal of Economic Behavior & Organization* 17 (1): 377–400. [136, 176]

Lighthall, Nichole R., Mara Mather, and Marissa A. Gorlick. 2009. "Acute stress increases sex differences in risk seeking in the balloon analogue risk task." *PLoS One* 4 (7): e6002. [149, 153, 156, 157]

List, John A. 2025. “The experimentalist looks within: Toward an understanding of within-subject experimental designs.” Working Paper 33456. National Bureau of Economic Research. [119, 120, 122, 151, 158, 159]

List, John A. 2026. *Experimental economics: Theory and practice.* University of Chicago Press. [120, 143, 156, 157]

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. “One swallow doesn’t make a summer: New evidence on anchoring effects.” *American Economic Review* 104 (1): 277–90. [143, 157]

Mather, Mara, Marissa A. Gorlick, and Nichole R. Lighthall. 2009. “To brake or accelerate when the light turns yellow? Stress reduces older adults’ risk taking in a driving game.” *Psychological Science* 20 (2): 174–76. [149, 153, 156, 157]

Matthews, Karen A., Brooks B. Gump, and Jane F. Owens. 2001. “Chronic stress influences cardiovascular and neuroendocrine responses during acute stress and recovery, especially in men.” *Health Psychology* 20 (6): 403. [178]

Murad, Zahra, Martin Sefton, and Chris Starmer. 2016. “How do risk attitudes affect measured confidence?” *Journal of Risk and Uncertainty* 52 (1): 21–46. [136, 176]

Nowacki, Jan, Hauke R. Heekeren, Christian E. Deuter, Jan David Joerißen, Aileen Schröder, Christian Otte, and Katja Wingenfeld. 2019. “Decision making in response to physiological and combined physiological and psychosocial stress.” *Behavioral Neuroscience* 133 (1): 59. [149, 153, 154]

Ockenfels, Margit C., Laura Porter, Joshua Smyth, Clemens Kirschbaum, Dirk H. Hellhammer, and Arthur A. Stone. 1995. “Effect of chronic stress associated with unemployment on salivary cortisol: Overall cortisol levels, diurnal rhythm, and acute stress reactivity.” *Psychosomatic Medicine* 57 (5): 460–67. [178]

Pabst, Stephan, Matthias Brand, and Oliver T. Wolf. 2013. “Stress effects on framed decisions: There are differences for gains and losses.” *Frontiers in Behavioral Neuroscience* 7: 142. [149, 153, 156, 157]

Parslow, Elle, and Julia Rose. 2022. “Stress and risk – Preferences versus noise.” *Judgment and Decision Making* 17 (4): 883–936. [118, 131, 142, 149, 153, 154]

Perugini, Marco, Marcello Gallucci, Fabio Presaghi, and Anna Paola Ercolani. 2003. “The personal norm of reciprocity.” *European Journal of Personality* 17 (4): 251–83. [171]

Porcelli, Anthony J., and Mauricio R. Delgado. 2009. “Acute stress modulates risk taking in financial decision making.” *Psychological Science* 20 (3): 278–83. [149, 153, 156, 157]

Rauh, Manfred. 2010. “Steroid measurement with LC–MS/MS. Application examples in pediatrics.” *Journal of Steroid Biochemistry and Molecular Biology* 121 (3-5): 520–27. [162]

Reynolds, Elizabeth K., Whitney M. Schreiber, Kathy Geisel, Laura MacPherson, Monique Ernst, and C. W. Lejuez. 2013. “Influence of social stress on risk-taking behavior in adolescents.” *Journal of Anxiety Disorders* 27 (3): 272–77. [149, 153, 155, 157]

Richter, David, Maria Metzing, Michael Weinhardt, and Jürgen Schupp. 2013. “SOEP scales manual.” SOEP Survey Papers 138. DIW Berlin. [171]

Schulz, P., and W. Schlotz. 1999. “The Trier Inventory for the Assessment of Chronic Stress (TICS): Scale construction, statistical testing, and validation of the scale work overload.” *Diagnostica* 45 (1): 8–19. [121, 127, 171, 179]

Shields, Grant S., Trey Malone, and Zach J. Gray. 2025. “Acute stress differentially influences risky decision-making processes by sex: A hierarchical bayesian analysis.” *Psychoneuroendocrinology* 172: 107259. [149, 153, 154, 157]

Smyth, Joshua M, Margit C Ockenfels, Amy A Gorin, Delwyn Catley, Laura S Porter, Clemes Kirschbaum, Dirk H Hellhammer, and Arthur A Stone. 1997. "Individual differences in the diurnal cycle of cortisol." *Psychoneuroendocrinology* 22 (2): 89–105. [130]

Sokol-Hessner, Peter, Candace M. Raio, Sarah P. Gottesman, Sandra F. Lackovic, and Elizabeth A. Phelps. 2016. "Acute stress does not affect risky monetary decision-making." *Neurobiology of Stress* 5: 19–25. [118, 149]

Stalder, Tobias, and Clemens Kirschbaum. 2012. "Analysis of cortisol in hair–state of the art and future directions." *Brain, Behavior, and Immunity* 26 (7): 1019–29. [121, 126, 127, 179]

Starcke, Katrin, and Matthias Brand. 2012. "Decision making under stress: A selective review." *Neuroscience & Biobehavioral Reviews* 36 (4): 1228–48. [117]

Staufenbiel, Sabine M., Brenda W.J.H. Penninx, Anne T. Spijker, Bernet M. Elzinga, and Elisabeth F.C. van Rossum. 2013. "Hair cortisol, stress exposure, and mental health in humans: A systematic review." *Psychoneuroendocrinology* 38 (8): 1220–35. [126, 179]

Tobler, Philippe N., Raymond J. Dolan, and Peter Dayan. 2009. "Risk-dependent reward value signal in human prefrontal cortex." *Proceedings of the National Academy of Sciences* 106 (17): 7185–90. [118]

Tobler, Philippe N., Wolfram Schultz, Quentin J. M. Huys, and Raymond J. Dolan. 2007. "Neural substrates of expected value and risk in decision making." *Neuron* 55 (3): 507–18. [118]

Trautmann, Stefan T. 2014. "Risk taking under stress: The role(s) of self-selection. A comment on Buckert et al.(2014)." *Frontiers in Neuroscience* 8: 197. [149]

Uhart, Magdalena, Rachel Y. Chong, Lynn Oswald, Ping-I Lin, and Gary S. Wand. 2006. "Gender differences in hypothalamic–pituitary–adrenal (HPA) axis reactivity." *Psychoneuroendocrinology* 31 (5): 642–52. [131]

Veszteg, Róbert F., Kaori Yamakawa, Tetsuya Matsubayashi, and Michiko Ueda. 2021. "Acute stress does not affect economic behavior in the experimental laboratory." *PLoS One* 16 (1): e0244881. [118, 142, 149]

Von Dawans, Bernadette, Clemens Kirschbaum, and Markus Heinrichs. 2011. "The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format." *Psychoneuroendocrinology* 36 (4): 514–22. [119, 121, 123, 126]

Von Helverson, Bettina, and Jorg Rieskamp. 2013. "Does the influence of stress on financial risk taking depend on the riskiness of the decision?" *Proceedings of the Annual Meeting of the Cognitive Science Society* 35 (35). [118, 149, 153, 155, 157]

Wennig, R. 2000. "Potential problems with the interpretation of hair analysis results." *Forensic Science International* 107 (1-3): 5–12. [126]

Yamakawa, Kaori, Hideki Ohira, Masahiro Matsunaga, and Tokiko Isowa. 2016. "Prolonged effects of acute stress on decision-making under risk: A human psychophysiological study." *Frontiers in Human Neuroscience* 10: 444. [118, 149, 153, 155, 157]

Appendix 3.A Literature review

Table 3.A.1. Literature summary

Paper	Mixed domains	Gain	Loss	N
Shields, Malone, and Gray (2025)	more risk-seeking	–	–	147
Parslow and Rose (2022)	no effect	–	–	194
Dreyer et al. (2022)	more risk-seeking (low-cortisol responders)	–	–	80
Veszteg et al. (2021)	no effect	–	–	192
Nowacki et al. (2019)	–	no effect	–	90
Kluen et al. (2017)	more risk-seeking (men); no effect (women)	–	–	103
Cahlíková and Cingl (2017)	–	more risk-averse (men); no effect (women)	–	146
Bendahan et al. (2017)[a]	–	more risk-seeking (directly after stressor); no effect (20 and 45 min after stressor)	–	352
Yamakawa et al. (2016)	–	more risk-averse	no effect	26
Sokol-Hessner et al. (2016) (within-subject)	no effect	no effect	–	120
Cueva et al. (2015) (within-subject)	more risk-seeking	–	–	34
Cohn et al. (2015)	more risk-averse	–	–	162
Finy et al. (2014)	more risk-seeking (low-constraint individuals); no effect (high-constraint individuals)	–	–	88
Gathmann et al. (2014)	no effect	–	–	38
Kandasamy et al. (2014) (within-subject[b])	–	more risk-averse (chronic); no effect (acute)	–	36
Buckert et al. (2014)	no effect	more risk-seeking (cortisol responders[c])	no effect	75
Daughters et al. (2013) (within-subject)	more risk-seeking (boys); no effect (girls)	–	–	132
Reynolds et al. (2013)	more risk-seeking (high social anxiety group); no effect (low social anxiety group)	–	–	34
Von Helverson and Rieskamp (2013) (within-subject)	more risk-seeking (low-risk); more risk-averse (high-risk)	–	–	70
Pabst, Brand, and Wolf (2013)	–	no effect	more risk-averse	80
Clark et al. (2012) (within-subject)	more risk-averse	no effect	more risk-averse	65
Mather, Gorlick, and Lighthall (2009)	more risk-averse (elderly); no effect (young adults)	–	–	85
Lighthall, Mather, and Gorlick (2009)	more risk-seeking (men); more risk-averse (women)	–	–	48
Porcelli and Delgado (2009)	–	more risk-averse	more risk-seeking	33

[a] Of these, approximately one-third underwent an experimental procedure to test the effect of stress on risk-taking immediately after the stressor, which is closest to our research question.
[b] The within-subject design applies to the study of chronic, not acute, stress.
[c] The authors note that limiting the analysis to cortisol responders makes the results correlational, which is also highlighted by Trautmann (2014).

Table 3.A.2. Within-subject change in standardized RP (alternative analysis)

	Change in Standardized RP		
	(Full)	(Gain)	(Loss)
RiskAverse	-0.026***	-0.022***	-0.031***
	(0.006)	(0.004)	(0.010)
RiskSeeking	0.031***	0.023*	0.039***
	(0.007)	(0.011)	(0.006)
Observations	3405	1721	1684
Observation-Level	Decision	Decision	Decision
R^2	0.022	0.015	0.030

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. Column (1) reports estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

In Table 3.A.3, we flip the analysis, keeping the same dependent variable, i.e., the dependent variable is the difference between the risk premia in *Stress* and *Control*, but changing the classification of risk aversion (or risk seeking), based on the decision in the *Stress* condition (rather than on decisions in the *Control* condition as in Table 3.A.2). If stress shifted individuals toward risk-neutrality, we would expect a positive coefficient for *RiskSeeking* and a negative coefficient for *RiskAverse*. However, we do not find this pattern, but rather a regression to the mean.

Table 3.A.3. Within-subject change in standardized RP (alternative analysis, flipped)

	Change in Standardized RP		
	(Full)	(Gain)	(Loss)
RiskAverse	0.030***	0.030***	0.029**
	(0.007)	(0.004)	(0.012)
RiskSeeking	-0.032***	-0.041***	-0.024***
	(0.007)	(0.012)	(0.006)
Observations	3405	1721	1684
Observation-Level	Decision	Decision	Decision
R^2	0.026	0.037	0.018

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. Column (1) reports estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

Appendix 3.B Informativeness of our null results

3.B.1 Statistical power in our design vs. a between-subject design

We evaluate statistical power under both our main within-subject design and our auxiliary between-subject analysis with the same total number of participants. The starting point is the standard decomposition (following the notation in List, 2025)

$$\overline{RP}_{it} \;=\; \pi_0 + \tau\, Stress_{it} + \mu_i + \varepsilon_{it}, \tag{3.B.1}$$

where $\overline{RP}_{it}$ is the average risk premium demanded by participant i at time t (averaged across all lottery choices in that session), $Stress_{it}$ indicates exposure to acute stress at time t, μ_i captures time-invariant subject-level heterogeneity in risk taking, and ε_{it} is an idiosyncratic error. Random assignment implies $Stress_{it}$ is independent of μ_i.

Simulation Model and Distributional Assumptions. To approximate power, we simulate datasets from a parametric data-generating process consistent with equation (3.B.1). We assume the participant-specific component is normally distributed, $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$, and idiosyncratic errors are normally distributed and homoscedastic, $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, independent across participants and independent of μ_i. We calibrate π_0 and the untreated-outcome variance σ_Y^2 using our *Control*-session data. Since we only observe one untreated outcome per participant, we cannot separately identify σ_μ^2 and σ_ε^2 in our data. We, therefore, impose a correlation parameter ρ for repeated outcomes within a participant and set $\sigma_\mu^2 = \rho\, \sigma_Y^2$ and $\sigma_\varepsilon^2 = (1-\rho)\sigma_Y^2$. Our baseline specification uses $\rho = 0.41$, based on the test–retest correlation of 0.41 risk-taking in a lottery choice task, reported in Falk et al. (2023). In their experiment, participants from the BonnEconLab subject pool who participated in two sessions of an experiment scheduled one week apart made a decision in a single incentivized lottery choice in both weeks. Since risk-taking in Falk et al. (2023) is measured using only lottery choices (versus 28 choices in our experiment), we view $\rho = 0.41$ as a conservative lower bound. Dohmen and Jagelka (2024) show that the test–retest correlation of personality trait measures and preference measures increases with the number of items. For example, the test-retest correlation of personality trait neuroticism increases from 0.59 when based on one item only to 0.85 when based on 12 items.

Effect Sizes from the Literature. Prior studies used different outcome measures. To make the effects across studies comparable, we translate each reported effect into Cohen's d (Table 3.B.1) and interpret d as a standardized treatment effect. For these power calculations, we use the studies from our literature review (compare Appendix 3.A).

Power Calculations. For each literature effect size d, we set the treatment effect in the simulations to $\tau = |d| \cdot \sigma_Y$. We then estimate τ using OLS under an equal-allocation between-subject design ($N = 140$) and participant fixed effects under a two-period cross-over within-subject design ($N = 140$), and test $H_0 : \tau = 0$ using a two-sided test at the conventional level of $\alpha = 0.05$. The power is the share of Monte Carlo replications for which the null is rejected.

Table 3.B.1. Power implied by effect sizes from prior studies

Study	Cohen's $\|d\|$	Power (BS, N=140)	Power (WS, N=140)
Shields, Malone, and Gray (2025)	0.350	0.539	0.965
Parslow and Rose (2022) – Detectable effect bound	0.470	0.792	0.999
Dreyer et al. (2022) – Responders vs. Control	0.348	0.532	0.962
Nowacki et al. (2019) – Women	0.935	1.000	1.000
Kluen et al. (2017) – Men	0.646	0.964	1.000
Cahlíková and Cingl (2017) – ATT Men	0.600	0.941	1.000
Bendahan et al. (2017) – Early (standard risk)	0.513	0.856	1.000
Yamakawa et al. (2016) – Lower bound	0.812	0.997	1.000
Cueva et al. (2015)	0.863	0.999	1.000
Cohn et al. (2015)	0.370	0.581	0.976
Finy et al. (2014) – Low-constraint adolescents	0.702	0.985	1.000
Buckert et al. (2014) – Responders vs. Control in gains	0.585	0.933	1.000
Daughters et al. (2013) – Boys	0.303	0.425	0.905
Reynolds et al. (2013) – High social anxiety adolescents	0.440	0.728	0.997
Von Helverson and Rieskamp (2013) – Interaction estimate	0.592	0.936	1.000
Pabst, Brand, and Wolf (2013) – Losses	0.672	0.976	1.000
Clark et al. (2012) – Losses	0.357	0.549	0.973
Mather, Gorlick, and Lighthall (2009) – Older adults	1.151	1.000	1.000
Lighthall, Mather, and Gorlick (2009) – Men	0.702	0.985	1.000
Lighthall, Mather, and Gorlick (2009) – Women	0.895	1.000	1.000
Porcelli and Delgado (2009) – Gains	0.450	0.754	0.998
Porcelli and Delgado (2009) – Losses	0.260	0.328	0.799
MDE Between-Subject	0.477	0.805	1.000
MDE Within-Subject	0.258	0.329	0.796

Notes: Power is computed by simulation from the data-generating process $Y_{it} = \pi_0 + \tau D_{it} + \mu_i + \varepsilon_{it}$ with $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$ and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. We calibrate π_0 and σ_Y^2 using participant-level *Control*-session means in our data and impose correlation $\rho = 0.41$ across repeated outcomes by setting $\sigma_\mu^2 = \rho\sigma_Y^2$ and $\sigma_\varepsilon^2 = (1-\rho)\sigma_Y^2$. For each study, we set$\tau = |d| \cdot \sigma_Y$ and estimate τ by OLS (between-subject) or by participant fixed effects (within-subject, $T = 2$, one treated period per participant; equivalent to estimating the first-differences equation). The final two rows calculate our study's minimum detectable effect size while maintaining at least 80% power, for our between-subject and within-subject analysis. Power is the fraction of simulated samples rejecting $H_0 : \tau = 0$ in a two-sided test at $\alpha = 0.05$ over 20,000 replications. *Derivations of Cohen's d.* Derivations of Cohen's d are reported in Appendix 3.B.2. In several studies, the authors find statistically significant effects at the conventional threshold of p-value < 0.05 on specific subsamples of their data. We indicate the subsamples chosen by the authors in the 'Study' column, but calculate our study's power to detect a similar or larger effect size in an ITT analysis based on our full sample. For brevity, we do not include Null results, because power under $d = 0$ is mechanically $\alpha = 0.05$. See Table 3.A.1 for a complete summary of the most related studies.

3.B.2 Derivations of effect sizes in Table 3.B.1

The entries below document how each Cohen's d in Table 3.B.1 is obtained from the original study's reports.

Shields, Malone, and Gray (2025). We base our power calculation on the frequentist robustness analysis reported in the Appendix. In the Type III Analysis of Deviance for risky-choice frequency, the Appendix reports a main stress effect, $\chi^2(1) = 5.43$, $p = .020$, and provides group summaries: stress-condition participants chose the risky option more frequently (mean proportion $= 0.58$, $SE = 0.02$) than control participants (mean proportion $= 0.52$, $SE = 0.02$), with $N_{\text{stress}} = 72$ and $N_{\text{control}} = 75$. Interpreting the reported SE values as standard errors of the group means across participants, we back out within-group standard deviations via $SD = SE\sqrt{n}$, compute the pooled standard deviation $SD_p = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}}$, and then compute Cohen's $d = (\bar{x}_{\text{stress}} - \bar{x}_{\text{control}})/SD_p \approx 0.35$ (sign defined as Stress—Control). We focus on this main-effect contrast (rather than stress×sex interactions) because, within their risky decision-making results, it is explicitly characterized as a small stress effect and is smaller than the prominent stress-by-sex and parameter effects reported for other decision components, making it a conservative benchmark for power calculations.

Parslow and Rose (2022) – Detectable effect bound. Not a point estimate. Included as a reference bound based on the paper's stated power: 90% power to detect approximately $d = 0.47$ at $\alpha = 0.05$ with N=194.

Dreyer et al. (2022). Computed from reported CD–AB cards means and SDs for the HighCort group (Treatment; $N_1 = 20$, mean= 5.17, SD= 8.42) versus the Non-Stress group (Control; $N_2 = 40$, mean= 2.15, SD= 8.81). Computed the pooled SD as $SD_p = \sqrt{\frac{(N_1-1)SD_1^2+(N_2-1)SD_2^2}{N_1+N_2-2}} \approx 8.68$, then Cohen's $d = (\bar{x}_1 - \bar{x}_2)/SD_p \approx 0.348$ (sign defined as Treatment—Control).

Nowacki et al. (2019) – Women. Approximated from the paper's figure (women-only subgroup; men insignificant). Read mean effects as Treatment $\approx$ 0.371 and Control ≈ 0.224, with error bars interpreted as standard errors ($\text{SE}_{\text{Trt}} \approx 0.043$, $\text{SE}_{\text{Ctrl}} \approx 0.038$) and $N_{\text{Trt}} = N_{\text{Ctrl}} = 15$. Converted SE to SD via $SD = SE\sqrt{n}$, yielding $SD_{\text{Trt}} \approx 0.167$ and $SD_{\text{Ctrl}} \approx 0.147$. Computed pooled SD as $SD_p = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}} \approx 0.157$, then Cohen's $d = (\bar{x}_{\text{Trt}} - \bar{x}_{\text{Ctrl}})/SD_p \approx 0.935$.

Kluen et al. (2017) – Men, Low Risk (pooled). Approximated from the relevant figure panel by reading the four cell means and assuming error bars are SEM with $n = 12$ per cell: Plac ≈ 14.5 (SEM ≈ 1.6), Yoh ≈ 12.2 (SEM ≈ 1.45), Cort ≈ 18.2 (SEM ≈ 2.4), and Cort+Yoh ≈ 16.9 (SEM ≈ 2.0). Converted SEM to within-cell SD via $SD = SEM\sqrt{n}$. Following the pooling implied by the authors' analysis, formed Control = Plac+Yoh and Treatment = Cort+(Cort+Yoh) with pooled means $\bar{x}_{\text{Ctrl}} = (\bar{x}_{\text{Plac}} + \bar{x}_{\text{Yoh}})/2$ and $\bar{x}_{\text{Trt}} = (\bar{x}_{\text{Cort}} + \bar{x}_{\text{Cort+Yoh}})/2$. Computed each pooled-group SD using the combined-sample variance identity (within-cell SS plus between-cell

SS around the pooled mean), then computed Cohen's d as $(\bar{x}_{\text{Trt}} - \bar{x}_{\text{Ctrl}})/SD_{\text{pooled}}$ using the pooled SD across the two pooled groups.

Cahlíková and Cingl (2017) – ATT Men. Used Cohen's d reported by the authors for the male subgroup simple comparison (treatment vs. control).

Bendahan et al. (2017) – Early (standard risk). Approximated from Figure 3 (solid gray bars). Read means as control ≈ 0.638 and stress ≈ 0.541. Assumed error bars are SEM, and allocation across time points is approximately equal, implying $n_{\text{control}} \approx 60$ and $n_{\text{stress}} \approx 58$. Converted SEM to SD via $SD = SEM\sqrt{n}$, then computed pooled-SD Cohen's d for (Stress—Control), yielding $d \approx -0.513$.

Yamakawa et al. (2016) – Lower bound. No SDs reported for the simple-effect contrast; computed the minimal $|d|$ consistent with a two-sided $p = 0.05$ test given $N_1 = 14$, $N_2 = 12$: $d_{\min} = t_{\text{crit}}\sqrt{1/n_1 + 1/n_2}$ with $df = 24$.

Cueva et al. (2015) (within-subject). Study reports a Wilcoxon signed-rank test, $Z = 2.17$. Converted to an effect-size correlation via $r = Z/\sqrt{N}$ using the within-subject sample size, and translated to a Cohen's-d metric using $d = 2r/\sqrt{1-r^2}$, yielding $d \approx 0.86$ (sign follows the reported direction of the effect).

Cohn et al. (2015). We recover Cohen's d from the reported regression coefficient and standard error in Table 1, Col. (1). The implied test statistic is $t = \hat{\beta}/\text{SE}(\hat{\beta})$ (here, $\hat{\beta} = -9.827$ and SE $= 4.177$, so $|t| \approx 2.35$). We then convert to a standardized mean difference using $d = |t|\sqrt{1/n_{\text{boom}} + 1/n_{\text{bust}}}$ with $n_{\text{boom}} = 85$ and $n_{\text{bust}} = 77$, yielding $|d| \approx 0.370$ (the sign follows the Bust—Boom coding).

Finy et al. (2014) – Low-constraint adolescents. Only the significance of the stress slope is reported for the low-constraint subsample (stress entered as a continuous predictor). Using total $N = 43$ and a two-sided $p = 0.03$ for the 1-df test of the slope implies residual degrees of freedom $df = N - 2 = 41$ and $|t| = t_{1-p/2,df} \approx 2.25$. We convert this 1-df test statistic to a partial correlation via $r = \sqrt{t^2/(t^2 + df)} \approx 0.332$, and map to a Cohen's-$d$ metric using $d = 2r/\sqrt{1-r^2} \approx 0.703$. This d reflects the standardized magnitude of the stress–risk-taking association (not a stress–control mean difference).

Buckert et al. (2014). Computed from reported risky-choice means and SDs for cortisol responders (N=26, mean=0.54, SD=0.26) vs. controls (N=20, mean=0.41, SD=0.16) using pooled SD.

Daughters et al. (2013) – Boys (within-subject). Study reports a 1-df simple-effect test for boys, $F(1, 58) = 4.5$. For a 1-df contrast, $t = \sqrt{F}$, so $t \approx 2.121$. With within-subject sample size $N = 58$, computed the paired standardized effect $d_z = t/\sqrt{N} \approx 0.279$. Converted d_z to an outcome-SD Cohen's d assuming within-subject correlation $\rho = 0.41$: $d = d_z\sqrt{2(1-\rho)}$, yielding $d \approx 0.303$.

Reynolds et al. (2013) – High social anxiety adolescents. Used the Cohen's d reported by the authors for the High SA subgroup, $d = -0.44$ (sign follows the paper's coding).

Von Helverson and Rieskamp (2013) – Interaction estimate. The paper's headline result is a 1-df interaction test, $F(1, 67) = 5.87$. To express the magnitude on a standardized scale comparable across studies, we convert the interaction F statistic to partial eta-squared, $\eta_p^2 = F/(F+df_2)$, then to Cohen's f via $f = \sqrt{\eta_p^2/(1-\eta_p^2)}$, and report the heuristic mapping $d \approx 2f$, yielding $d \approx 0.592$. This d summarizes the standardized magnitude of the interaction (not a stress–control mean difference).

Pabst, Brand, and Wolf (2013). Used the reported simple-effect test for the loss-domain stress–control contrast: $F(1,76) = 4.52$. For a 1-df contrast, $t = \sqrt{F}$. Converted to Cohen's d via $d = t\sqrt{1/n_1 + 1/n_2}$ with $N_1 = N_2 = 20$, yielding $d \approx 0.672$ (sign defined as Stress—Control).

Clark et al. (2012) – Losses. Within-subject design. Reported losses-only contrast $t(56) = 2.48$ implying $d_z = t/\sqrt{n}$ with $n = 57$. Converted d_z to an outcome-SD Cohen's d assuming $\rho = 0.41$: $d = d_z\sqrt{2(1-\rho)}$, yielding $|d| \approx 0.357$ (sign depends on coding threat—safe).

Mather, Gorlick, and Lighthall (2009) – Older adults. Converted reported independent-groups t-test for older adults: $t(38) = 3.64$. Older-adult sample totals N=40 implies approximately 20/20 split across stress/control. Computed $d = t\sqrt{1/n_1 + 1/n_2}$ yielding $|d| \approx 1.151$ (sign depends on Stress—Control).

Lighthall, Mather, and Gorlick (2009) – Men. Backed out within-cell SDs from reported 95% CI half-widths and N, then computed Cohen's d as the stress–control mean difference divided by the pooled SD.

Lighthall, Mather, and Gorlick (2009) – Women. Backed out within-cell SDs from reported 95% CI half-widths and N, then computed Cohen's d as the stress–control mean difference divided by the pooled SD.

Porcelli and Delgado (2009) – Gains. Used the Cohen's d reported by the authors for the stress effect in the gain domain.

Porcelli and Delgado (2009) – Losses. Used the Cohen's d reported by the authors for the stress effect in the loss domain.

3.B.3 What can we learn from our null result?

The above power calculations demonstrate that our study would have been powered to detect the results of previous studies with at least 80% power. Additionally, our minimum detectable effect (MDE) size is Cohen's $d = 0.477$ in a between-subject analysis and $d = 0.258$ in a within-subject analysis with power 80% at $\alpha = 0.05$.

Based on our null result, we compute a post-study probability (PSP) of a relevant relationship between acute stress and risk-taking behavior. We follow List (2026) and calculate the post-study probability of a true association after a null result as *one minus* the fraction of the non-true associations being declared non-true and all associations that are declared non-true:

$$PSP_{Null} = \frac{(1-\alpha)(1-\pi)}{\beta\pi + (1-\alpha)(1-\pi)}, \tag{3.B.2}$$

where $(1-\alpha)(1-\pi)$ represents the probability of a non-true association being declared non-true, and $\beta\pi$ represents the probability of a true association being declared non-true. We set $\beta = 0.20$ and $\alpha = 0.05$. To find a reasonable prior π, we calculate the proportion of statistically significant associations reported in the studies presented in our literature summary in Table 3.A.1 in Appendix 3.A. We follow the suggestions in Maniadis, Tufano, and List (2014) and consider a range of possible priors. On the lower bound, we only consider studies that report a statistically significant result for their entire sample. We count 6 out of 24 papers as reporting a statistically significant association,[51] resulting in $\pi = 0.25$. On the upper bound, we consider all studies that report a significant association between *acute* stress and risk-taking at least in one subsample. We count 18 out of 24 papers as reporting a statistically significant association,[52] resulting in $\pi = 0.75$. For this range of priors, our well-powered null result lowers the likelihood of a causal effect of acute stress on risk taking by 18.4 to 36.3 percentage points: with a prior of 25%, a null result at 80% power implies an update to a PSP of 6.6%, while a prior of 75% implies a PSP of 38.7%. In both cases, this range constitutes a large update from the prior, highlighting that there is considerable information in our null result (see also Maniadis, Tufano, and List, 2014; Abadie, 2020; List, 2026).

51. Lighthall, Mather, and Gorlick (2009), Porcelli and Delgado (2009), Von Helverson and Rieskamp (2013), Cohn et al. (2015), Cueva et al. (2015), and Shields, Malone, and Gray (2025)

52. Lighthall, Mather, and Gorlick (2009), Mather, Gorlick, and Lighthall (2009), Porcelli and Delgado (2009), Clark et al. (2012), Daughters et al. (2013), Pabst, Brand, and Wolf (2013), Reynolds et al. (2013), Von Helverson and Rieskamp (2013), Buckert et al. (2014), Finy et al. (2014), Cohn et al. (2015), Cueva et al. (2015), Yamakawa et al. (2016), Bendahan et al. (2017), Cahlíková and Cingl (2017), Kluen et al. (2017), Dreyer et al. (2022), and Shields, Malone, and Gray (2025)

Appendix 3.C Experiment

3.C.1 Assumption necessary for within-subject designs

List (2025) stresses that, in addition to assumptions necessary to recover an average treatment effect in a between-subject design, a within-subject design relies on three additional assumptions: balanced panel, temporal stability, and causal transience. Below, we argue that each of these assumptions is met in our context.

3.C.1.1 Balanced panel

As noted by List (2025), this assumption is typically not a concern in laboratory experiments. Moreover, our incentive structure is in line with his recommendation to backload incentives, as participants received their payout at the end of the second session. Of the 163 participants who showed up for their first experiment session, 156 completed both experiment sessions. Of the seven participants who did not complete both sessions, two participants (one in the *Stress* and one in the *Control* condition) chose to leave the experiment during the first session. This type of attrition would have also occurred had we conducted a between-subject design. Of the remaining five participants, three excused themselves from the second session due to illness or other appointments. Two of these had experienced the *Stress* condition as their first session, while one had experienced the *Control* session as their first session. Two participants (both of whom had experienced the *Stress* condition as their first session) did not show up for the second session without a statement of reasons.

In our analysis, we only include participants who completed both sessions, i.e., we lose 3% of participants due to our within-subject design. We consider it unlikely that experiencing the first experimental session induced selective attrition for the second experimental session, and that this low attrition may have affected our results.

3.C.1.2 Temporal stability

For the temporal stability assumption to be violated, participants' risk-taking behavior would need to change systematically from one week to the next, when elicited on the same day of the week, at the same time, and under the same conditions as the previous elicitation, and conditional on treatment condition. We consider such a change as highly unlikely. Nevertheless, note that possible temporal effects would be averaged out in our effect analysis since we counterbalance the order in which participants experience the *Control* and *Stress* sessions.

3.C.1.3 Causal transience

This assumption implies the absence of 'carryover' effects from previous treatments and rules out anticipation effects whereby future treatments influence current behavior. Violation of the causal transience assumption would imply that we conduct what List (2025) denotes an 'overt' within-subject design. Such a violation could arise, for example:

- if acute stress experienced by participants assigned to the *Stress* condition in their first-week session persisted into their second-week experiment session, when they are in the *Control* condition. This could lead to an underestimation of effect sizes. However, such persistence appears implausible given the short-lived nature of acute stress.
- Alternatively, participants who experience the *Control* condition in their first-week session might respond more strongly to stress in their second-week *Stress* session. While we do not find evidence that the physiological stress response differs depending on whether stress is administered in the first or second week (see Appendix 3.D.1), such an effect, if present, could lead to an overestimation of effect sizes.

While we consider these violations unlikely, our design allows us to assess their potential magnitude. Specifically, we report our main results separately for participants who experienced the *Control* condition in their first session and those who experienced the *Stress* condition first (Appendix 3.D.4). Because treatment order differs across participants, potential violations of causal transience would affect these groups asymmetrically. In the examples above, persistence effects would only affect participants assigned to *Stress* first, whereas differential stress responsiveness would only affect those assigned to *Control* first. Regardless of treatment order, we find no evidence of an effect of acute stress on risk-taking, with all p-values well above conventional significance levels.

3.C.2 Biomarker measurement: Overview

We collected saliva samples to measure salivary cortisol throughout the experiment. Our cortisol elicitation protocol was designed to address several well-known challenges associated with measuring salivary cortisol in laboratory experiments. First, salivary cortisol does not rise immediately after a psychological stressor. Instead, cortisol is released into the bloodstream with a delay of approximately 5–20 minutes, and biologically active (free) cortisol appears in saliva within 2–3 minutes (Bozovic, Racic, and Ivkovic, 2013). As a consequence, salivary cortisol peaks around 20–30 minutes after the onset of a stressor and typically returns toward baseline within 40–60 minutes. Second, cortisol levels follow a strong daily rhythm: They peak roughly 30 minutes after awakening and decline steadily throughout the day. In the afternoon, the diurnal slope is relatively flat, reducing the risk of confounding experimental effects with natural fluctuations. Third, participants may experience elevated cortisol levels upon arrival due to uncertainty and novelty, regardless of being assigned to the *Stress* or *Control* condition (Dohmen, Rohde, and Stolp, 2023). If not accounted for, such initial arousal could mask treatment-induced effects.

To capture meaningful cortisol responses while avoiding confounds, we incorporated several design elements: (1) Repeated sampling throughout the experiment to track the cortisol trajectory across the entire window in which stress responses occur. (2) Afternoon-only sessions with a consistent starting time to minimize diurnal variation. (3) A one-hour adaptation period between arrival and treatment onset to allow any arrival-related cortisol elevation to subside.

In addition to salivary cortisol, we also measured two additional biomarkers: (1) participants' heart rate, and (2) participants' hair cortisol. Participants were free to withdraw from the experiment at any time, while still receiving compensation for the time they had already invested in the experiment. After the data collection was completed, we debriefed study participants via email on the purpose of the *Stress* protocol and biomarker measurement, and offered our contact data for any further questions. The role of salivary cortisol as a biomarker of acute stress, as well as procedures for saliva sampling, heart rate measurement, and hair sampling, are described in more detail below.

3.C.2.1 Salivary cortisol as a biomarker for the psychological stress reaction

Salivary cortisol is a biomarker for activation of the hypothalamic–pituitary–adrenocortical (HPA) axis. The HPA axis is one of the body's major neuroendocrine systems, involving interactions between the hypothalamus, pituitary gland, and adrenal glands. It regulates digestion, energy balance, immune function, mood, and emotional processes, and is "vital for supporting normal physiological functions and regulating other systems" (Dickerson and Kemeny, 2004, p. 356).

A stress reaction begins with the release of adrenocorticotropic hormone (ACTH) by the anterior pituitary, which triggers the secretion of cortisol from the adrenal cortex into the bloodstream. Free cortisol appears in saliva within a few minutes, making saliva a reliable medium for detecting biologically active cortisol (Kirschbaum and Hellhammer, 1994).

3.C.2.2 Saliva sampling procedures

We collected saliva samples to measure salivary cortisol throughout the experiment. Participants were informed in the invitation email and during the study's general information that saliva samples would be collected. At the beginning of the session in the first week, participants received detailed instructions on saliva sampling. An experimenter demonstrated how to handle the Salivette®. The experimenter informed participants that they should not touch the cotton swab of the Salivette® by hand

and should try to chew on it for approximately 2 minutes. A leaflet detailing the most important steps was distributed to all cubicles.

During the experiment and before the first saliva sample, participants were instructed on-screen that several saliva samples would be taken during the experiment. Furthermore, it informed participants that they could read the instructions for sampling on the leaflet again in their cubicle, and that each sample would be announced via *(i)* a screen notification and *(ii)* the sound of a triangle (the musical instrument). The Salivette® for the first saliva sample was distributed to the cubicles before the onset of the experiment. Subsequently, a new Salivette® was handed out to the participants when the previous one was collected. Thereby, the total number of saliva samples and their timing remained obscured from the participants. Each saliva sample was accompanied by a screen that briefly reminded participants to try keeping the cotton swab for 120 seconds in their mouth and integrated a timer to facilitate timekeeping.

Taking a saliva sample took approximately three minutes: The participants first chewed on a cotton swab of a Salivette® for two minutes, and it then took approximately one minute for the experimenters to collect the used Salivettes® and distribute the Salivettes® and stickers for the next saliva sample.

To avoid compromised samples due to glucose intake, participants were not allowed to eat during the experiment. To prevent diluted samples, we also asked participants not to drink anything except for water that we offered them during scheduled water breaks.[53]

Saliva sample number one was taken 30 minutes after participants' arrival in the laboratory, which is referenced as minute -33 in Figure 3.1 in Section 3.2. At this point, cortisol might be elevated due to potential stress and arousal upon entering the laboratory, as well as any pre-laboratory stress (Dohmen, Rohde, and Stolp, 2023) warranting an adaptation period as implemented in our design. Sample number two, our baseline sample, was taken around 30 minutes later, shortly before the onset of the experimental treatment. The cortisol levels are very similar between the the treatment and control conditions and do not differ statistically (two-sided t-test, p-value $= 0.922$). Samples three to five allow us to measure participants' responses to the treatment conditions and their cortisol levels during the risk elicitation.

The saliva samples obtained during the experiment sessions were stored in a freezer (-20) and sent on dry ice to the Dresden LabService GmbH, where, after thawing, the samples were centrifuged for 5 minutes at 3000 rpm, which resulted in a clear supernatant of low viscosity, to determine salivary cortisol. Salivary concentrations were measured using a commercially available chemiluminescence immunoassay with high sensitivity (Tecan - IBL International, Hamburg, Germany; cat-

53. We scheduled these in a way that ensured sufficient distance in timing to the next saliva sample.

alogue number R62111). The intra- and interassay coefficients of variance were below 9%.

3.C.2.3 Heart rate measurement

To measure participants' heart rate, we equipped them with a pulse sensor (Polar® Verity Sense), which is worn on the forearm or upper arm during the entire experiment. At the beginning of each session, participants were instructed on the use of the pulse sensor via on-screen information. An experimenter then controlled the correct fit and application for each participant. The operation of the pulse sensors did not require any qualified medical personnel.

3.C.2.4 Hair sampling procedures

We also collected hair samples from participants who were willing to provide such a sample. The hair samples were stored in aluminum foil and also sent to the Dresden LabService GmbH for analysis. They were analyzed using a Liquid chromatography-tandem mass spectrometry (LC–MS/MS)-based method to determine the individual level of cortisol concentration in the hair (Rauh, 2010; Gao et al., 2013). The hair samples were washed with isopropanol, and cortisol was extracted from 10 mg whole, nonpulverized hair by methanol incubation. A column switching strategy for online solid phase extraction (SPE) was applied, followed by analyte detection on an SCIEX Triple Quad 6500 mass spectrometer. The intra- and interassay coefficients of variation were 9.1% and 12.7%. The average weight for segment 1 was 7.32 mg (SD: 0.73), and 7.21 mg (SD: 0.96) for segment 2.

3.C.3 Chronological sequence of study

The experiment was conducted at the University of Bonn's Laboratory for Experimental Economics (BonnEconLab) between June and November 2022. We additionally conducted a pilot study with 12 participants in May 2022 to test the experimental procedures. Each session took approximately two and a half hours. During the 'Adaptation Period', we conducted all registration, provided study information, and instruction procedures as described below. During the 'Treatment Phase', participants underwent either the *Stress* or the *Control* condition. The treatment was followed by a set of questionnaires (compare Appendix 3.C.9), and a voluntary hair sample as well as payout in participants' second session.

3.C.3.1 Adaptation phase

The experiment took place in two rooms. Participants entered Room A, the room that hosts 24 computer cubicles. In the experiment, we only used 8 of these cubicles and randomly assigned the numbers 1 to 8 to them. In the middle of the room, we put a desk with two chairs, where the committee was placed during the *Stress* treatment. Adjacent to Room A is Room B, where participants received instructions and prepared for the tasks.

On arrival, participants entered Room A, registered, and then were directed to Room B. In the first experiment session, once all participants had arrived, the experimenters familiarized the participants with the consent form and general information about the study, stressing the handling of sensitive physical data such as saliva samples. Participants were explicitly informed that they would only participate in the experiment if they provided consent for participation. After participants had signed the consent form, an experimenter demonstrated saliva sampling to participants. Participants were also informed that they could also provide a hair sample for an additional payment at the end of the second session. Before participants returned to Room A, they picked a random card with their cubicle number and got seated in their cubicle. In their second session, they were guided to the same cubicle they had used in their first session of the experiment.

Each cubicle was equipped with a computer, magazines, the Salivette® for the first saliva sample, a printout reminding participants how to retrieve a saliva sample, a sticker, and a pen to label the sticker for the first saliva sample. At their cubicle, participants put on their heart rate monitoring devices, whose fit and function were then checked by one of the experimenters. At the start of the experiment, participants received instructions on the screen and familiarized themselves with the risk elicitation procedure by completing six multiple price lists as unincentivized practice questions. They could then read magazines and relax in the remaining time before the onset of the treatment. They were allowed to go to the bathroom[54] and drink water after the first saliva sample (minute: -33).

3.C.3.2 Treatment phase

Approximately one hour after arrival, the participants were prompted to take their second saliva sample and then given on-screen instructions for either the *Stress* or the *Control* procedure.

***Stress* condition.** In the *Stress* condition, participants were informed that they would undergo three challenges and that they had to move to Room B for further instructions. The details of the TSST-G procedures are detailed in Appendix 3.C.4.

During this time, they were allowed to drink some water. While the participants prepared themselves for the first challenge, three experimenters set up Room A, bringing in and turning on the cameras, inviting the committee to the room, and preparing the committee desk by placing a timer, blank sheets of paper, pens, and

54. Participants were free to take a bathroom break at other points in the experiment, but experimenters explicitly announced bathroom breaks before and after the challenges/tasks to avoid disruption during the TSST-G as much as possible.

a set of instructions on it. They did this silently, so that the participants in Room B were not aware of a change happening.

As the participants were in Room B during this time, they only took account of the newly set-up cameras and the committee when they re-entered Room A after the preparation time had ended. After the first challenge was over (see Appendix 3.C.4 for details), the committee reminded the participants that they would see each other again and then moved to Room B. The participants sat down in their cubicles, closed their curtains, and completed the first risk elicitation block. Afterwards, they were prompted to take the third saliva sample.

After providing the third saliva sample, participants received instructions on their screen to stand up and turn around again for the second challenge. The committee then re-entered Room A and gave instructions on the second challenge. After the second challenge was over, the procedure from above was repeated: the committee reminded the participants that they would see each other again and left the room, the participants sat down, closed their cubicle curtains, proceeded with the second risk elicitation block, and then provided saliva sample four.

After taking the fourth saliva sample, participants were instructed to stand up and turn around again for the last challenge. Upon completion, the committee left the laboratory. The participants sat down, closed their curtains, and provided the last saliva sample. Meanwhile, an experimenter turned off and removed the cameras. The experimenters offered water and announced a bathroom break.

***Control* condition.** In the *Control* condition, the sequence of events was equivalent except for the following differences: first, participants prepared for the first task directly in their cubicles instead of going to room B. Second, in contrast to the *Stress* condition, no video cameras were installed in Room A. Third, the tasks were conducted not by a designated committee, but by two of the experimenters, who did not wear laboratory coats and whom the participants may have already seen in other parts of the experiment. They sat down behind the desk in the middle of room B and brought their materials and a watch.

As in the *Stress* condition, participants complete the risk elicitation blocks between the tasks and take the same number of saliva samples at the exact same time as in the *Stress* condition.

After the TSST-G block was over, the participants first self-assessed how stressed they felt during the challenges/tasks and subsequently filled out several questionnaires, which differed by session condition (compare, Appendix 3.C.9). In the second week, they were also asked if they would provide a hair sample for an additional payment. The participants who agreed to do so were paid the extra amount during payout and went to Room B to have their hair sample taken by one of the experimenters.

3.C.4 The Trier Social Stress Test for Groups (TSST-G)

3.C.4.1 Stress induction in the *Stress* condition

In the *Stress* condition, a two-person committee (one male, one female) wearing laboratory coats observed participants.[55] In addition, we informed participants that their performance would be recorded on video. We made slight modifications to the standard Trier Social Stress Test for Groups (TSST-G) protocol to align with standards regarding potential deception in experimental economics and to improve procedural control. First, to adhere to the non-deception rule of the BonnEconLab, we excluded elements of the original protocol that could be misleading. For example, the original protocol introduces committee members as 'experts in human expressions'. Since our committee lacked such training, we omitted this description. Second, to enhance experimental control, we presented participants with a list of predefined job options for the mock interview, whereas the original protocol required them to come up with a job themselves. We randomized between two lists, A and B, of predefined job options. List A – video game developer, digitalization officer, fundraiser, research group leader, journalist, life coach, sustainability officer, airplane pilot. List B – ambassador, decarbonization manager, event manager, career coach, communications manager, head of emergency services, helicopter pilot, project manager.

Third, to improve efficiency, we allowed for group sizes of up to eight participants, compared to six in the original TSST-G. Lastly, we added a third challenge, reciting the alphabet backward (compare, Cingl, 2019), to extend the stress period and allow for two phases of risk elicitation. The third challenge ensured that participants were stressed throughout the entire risk-elicitation period, i.e., once the second challenge was completed, participants would still experience anticipatory stress in expectation of the third challenge. While all participants began with the mock interview, we randomized the order of the arithmetic and the alphabet challenges across sessions.

During each challenge, participants stood in a row (segregated by cubicle walls) facing the committee. The committee called them individually to step forward to perform the challenge. Participants could be summoned multiple times within one challenge. Since the committee selected participants in random order, they were unaware of when or how often they would be called upon. The committee's role was to manage turn-taking, monitor timing, prompt participants to continue speaking, ask questions, and instruct participants to restart in the event of a mistake. The committee members were trained to maintain a neutral demeanor throughout, avoid any emotional responses, empathy, interest, or verbal feedback.

At the beginning of the treatment phase in the *Stress* condition, participants saw an on-screen announcement which informed them of the three upcoming challenges.

55. The committee members did not assume any other roles during the experiment.

They were then escorted by an experimenter to a separate room, where they received verbal and written instructions for the first challenge, a mock job interview. These instructions informed participants that two additional challenges would follow, with further information provided in due time. After reading the instructions, we granted participants three minutes to prepare for the mock interview. The instructions explicitly stated that note-taking was permitted during preparation but that these notes could not be used during the interview. They also informed that each interview would last two minutes and would not focus on specific qualifications but on participants' personal aptitude for the job.

Following the preparation period, participants returned to the main experiment room. During their absence, the room was rearranged to accommodate the committee: cameras were set up and activated, and a desk for the committee was prepared. As participants re-entered, the committee was already seated, and the first challenge began immediately.

At the end of the first challenge, the committee reminded participants that a second challenge would follow. The committee then left the room, and participants completed the first set of risk elicitation tasks. Subsequently, the committee returned to introduce the second challenge. After this challenge, and once the committee had exited again, participants completed the second set of risk elicitation questions, followed by a third challenge.

Importantly, we consistently referred to these components as 'challenges' in the *Stress* condition to distinguish them from the control protocol, where they were referred to as 'tasks'. The written and oral instructions provided to participants in the *Stress* conditions are in Appendix 3.C.5.

3.C.4.2 The *Control* condition

Prior to the intervention phase in the *Control* condition, participants received an on-screen message explaining that three tasks would follow. The message also provided procedural details and explicitly stated that neither the participants nor their performance would be evaluated or judged. Participants were informed that the task would be to read a text that they would then receive and that they would have six minutes to prepare for this task.[56] The experimenters and the participants read the text out loud for the same duration as the mock interview challenge in the *Stress* condition. Instead of the difficult arithmetic challenge, participants counted from one to 100 together with the experimenters, and instead of the difficult alphabet challenge, they recited the alphabet together. None of these tasks needed to be completed simultaneously with the other participants or the experimenters.

56. By allowing for a preparation time of six minutes, we ensured that the timing of each task and risk elicitation segment was identical across both the *Stress* and *Control* conditions, maintaining temporal consistency throughout the experiment

The written and oral instructions provided in the *Control* condition is available in Appendix 3.C.6.

3.C.5 Translation of the instructions: *Stress* condition

Please see Appendix 3.F for the on-screen instructions of the entire experiment.

3.C.5.1 Instructions challenges: Read aloud by the experimenter

Please take a seat and read the instructions on the table for the following challenges. Please do not talk to the other participants.
[The participants sit and read the instructions. After their preparation time is over, the experimenter says:]
I will now take you back to the computer room. Before you follow me into the room, I will briefly explain the procedure again: Each of you will have about two minutes to present yourselves to a committee. You will be called up by your cubicle number. You will be filmed during your presentation. The committee members will closely monitor your behavior and take notes. The committee may ask you follow-up questions, inquiries, or questions not directly related to your presentation, including questions directed at you during presentations by other applicants.
The other two challenges will be explained to you by the committee in due course. You will be called at random and can be called again at any time. Do you have any questions?

3.C.5.2 Instructions challenges: Read by the participants

The first challenge is as follows: Imagine you have applied for one of the positions listed below and are now at the interview. In contrast to a real interview, you have to give an approximately 2-minute presentation in which you convince a committee that you are the best candidate for the position.
You can choose from the following positions:
computer game developer, digitization officer, fundraiser, research group leader, journalist, life coach, sustainability officer, pilot (aircraft).

OR [list of eight jobs randomized between sessions]

ambassador, decarbonization manager, event manager, career coach, communications manager, head of emergency services, pilot (helicopter), project manager.
You should focus primarily on your personal strengths – in other words, your personal characteristics that differentiate you from the other candidates for this position. You shouldn't focus on your knowledge and skills – assume the committee already knows about your academic and professional trajectory. The committee members will closely monitor your behavior and take notes. You will also be filmed. You

should try to make the best possible impression and assume the role of the applicant as best you can for the duration of the challenge. The committee may ask you follow-up questions, inquiries, or questions not directly related to your presentation, including questions directed at you during presentations by other applicants.
Please maintain an upright posture during all challenges and do not put your hands in your pockets.
The other two challenges will be explained to you by the committee in due time. You will be called up randomly and can be called up again at any time. The challenges will last approximately one hour in total. Please remember your cubicle number, as this will be used to call you up. Once you've read these instructions, you have 3 minutes to prepare for the first challenge. You can take notes, but you are not allowed to take them with you to the presentation.

Please give a show of hands if you have any questions. Please do not chat with the other participants.

3.C.5.3 Instructions given by the committee

Challenge 1. Number X: Please step forward. Please begin.
[The interview takes place. The participants can be interrupted. They can be asked questions. They can be asked to sit down before their time is up. They can be called up again.]
Please return to your seat and close the curtain. You may continue working on the computer.

Challenge 2 (or 3). The next challenge is to recite the alphabet backwards, starting at a certain letter. Please say it out loud. For example, if you get the letter M, the next letter is L. Please say it as quickly and correctly as possible. If you make a mistake, either you or another participant will have to try again or start from the beginning. You may also be asked to repeat the challenge or continue for another participant. We will call you by your cubicle number. You can also be called several times. Number X: Please step forward. Please begin with the letter ...

Challenge 2 (or 3). The next challenge is to solve math problems. Please count backward from the number we give you to zero in steps of 17. For example, if you are given the number 102, you should calculate 102 - 85 - 68 - 51 and so on. Please calculate as quickly and correctly as possible. If you miscalculate, either you or another participant will have to try again or start from the beginning. You may also be asked to solve the challenge repeatedly or to continue it for another participant. We will call you with your cubicle number. You can also be called several times. Number X: Please step forward. Please begin with the number ...

3.C.6 Translation of the instructions: *Control* instructions

Please see Appendix 3.F for the on-screen instructions of the entire experiment.

3.C.6.1 Instructions given by the committee

Task 1. Please place your drinking cups on the table. We will now start with the first task. This consists of reading aloud together the text that was handed out to you earlier. We will now read aloud together for 16 minutes. You only need to speak loud enough for your voice to be audible for all other participants and the staff to hear you. If you finish the task prematurely, please start again.

Task 2 (or 3). The second/third task is to recite the alphabet together with the other participants and staff for 4 minutes. You can take short breaks and recite the alphabet at your own pace. You only need to speak loud enough for your voice to be audible for all the other participants and the staff to hear you. If you finish the task prematurely, please start again.

Task 2 (or 3). The second/third task is to count to 100 (1, 2, 3, 4, ... 100) together with the other participants and the staff for 4 minutes. You can take short breaks and count at your own pace. You only need to speak loud enough for your voice to be audible for all other participants and the staff to hear you. If you finish the task prematurely, please start again.

3.C.7 Details on risk-taking decisions

Between the challenges/tasks of the TSST-G procedure, we elicited participants' risk-taking behavior via 28 lotteries. The lotteries were presented between the tasks in blocks of 14 each. We informed participants that one of the lotteries would be randomly selected for payment in each session. From the selected lottery, one row from the corresponding multiple price list, displaying a choice between a lottery and a safe payment, was randomly drawn. If the participant had chosen the lottery in that row, the lottery was played out, and the participant was paid accordingly. If the participant had selected the safe payment, the respective amount (gain or loss) was added to their experimental endowment of 535 experimental points (exchange rate of point to EUR 15:1). The endowment was high enough to ensure that participants could not incur financial losses from participating in the experiment.

During the first phase of each session, participants received detailed information on their financial incentives to respond accurately in the lottery decisions, and had to pass a comprehension check to ensure understanding. They then completed six non-incentivized decisions (three in the gain and three in the loss domain) to familiarize themselves with this type of decision.

Our set of lotteries is adapted from Bruhin, Fehr-Duda, and Epper (2010). Specifically, we use the set of lotteries used in the Beijing 2005 experiment. Table 3.C.1 provides an overview of all used MPLs, including the probabilities of the lottery, p1 and p2, as well as the highest and lowest safe payments, x1 and x2. All lotteries are visible in the screen shots in Appendix 3.F.

Table 3.C.1. Overview of lotteries used in the study

p1	x1	x2	p1	x1	x2
0.1	20	10	0.1	-10	-20
0.5	20	10	0.5	-10	-20
0.9	20	10	0.9	-10	-20
0.05	40	10	0.05	-10	-40
0.25	40	10	0.25	-10	-40
0.5	40	10	0.5	-10	-40
0.75	40	10	0.75	-10	-40
0.95	40	10	0.95	-10	-40
0.05	50	20	0.05	-20	-50
0.25	50	20	0.25	-20	-50
0.5	50	20	0.5	-20	-50
0.75	50	20	0.75	-20	-50
0.95	50	20	0.95	-20	-50
0.05	150	50	0.05	-50	-150

3.C.8 Recruitment criteria and behavioral rules

The following individuals were not allowed to participate in the study according to the invitation email:

- currently pregnant or breastfeeding
- somatic, neurological or psychiatric illnesses
- heavy smoking
- addiction to alcohol or drugs
- a disease or impairment of the heart
- a disease or impairment of the kidneys
- an eating disorder
- a metabolic disorder
- diabetes
- severe underweight or overweight
- weakened immune system on the day of the experiment
- diarrhea on the day of the experiment
- painkillers on the day of the experiment

Furthermore, the invitation email specified the following behavioral rules on the day of the experiment:

- no consumption of alcoholic beverages in the last 12 hours before the experiment
- no sports in the last 12 hours before the experiment
- no visit to the dentist in the last 12 hours before the experiment
- no meals two hours before the experiment
- no teeth brushing one hour before the experiment
- no consumption of sugary, carbonated or caffeinated drinks one hour before the experiment
- no smoking one hour before the start of the experiment

3.C.9 Overview questionnaires

3.C.9.1 Participants' first session

- Self-assessed stress, effort, exhaustion, calmness, tension, satisfaction, and tiredness (Dohmen, Rohde, and Stolp, 2023)
- Altruism (Falk et al., 2018)
- Reciprocity (Perugini et al., 2003; Fehr and Schmidt, 2006) according to SOEP (Dohmen et al., 2008, 2009; Richter et al., 2013)
- General and domain specific risk (Dohmen et al., 2011) according to SOEP (Richter et al., 2013)
- Trier Inventory for the Assessment of Chronic Stress (TICS, Schulz and Schlotz, 1999)
- Demographics: age, gender, year of high school degree, last mathematics grade in school, field of study, disposable monthly income (after rent)
- Well-being on session day and observation of behavioral rules (sickness today, alcohol consumed in last 12 hours, sports today, dentist visit today, teeth brushed an hour before the session, food intake two hours before the session, consumption of carbonated, sugary, or caffeinated drinks one hour before the session, smoking one hour before the session, painkiller intake today)

3.C.9.2 Participants' second session

- Self-assessed stress, effort, exhaustion, calmness, tension, satisfaction, and tiredness (Dohmen, Rohde, and Stolp, 2023)
- 15-item Big-Five (Costa and McCrae, 1985) according to SOEP (Gerlitz and Schupp, 2005)
- Experiment-related questions: knowledge about TSST-G, knowledge about multiple price lists, number of previous experiments at BonnEconLab, acquaintance with other participants in the session
- Health-related questions: size in cm, weight in kg, self-assessed general health state, regular intake of cortisone medication, previous or current mental condition, use of hormonal contraception

- Well-being on session day and observation of behavioral rules (sickness today, alcohol consumed in last 12 hours, sports today, dentist visit today, teeth brushed an hour before the session, food intake two hours before the session, consumption of carbonated, sugary, or caffeinated drinks one hour before the session, smoking one hour before the session, painkiller intake today)

Appendix 3.D Additional results

3.D.1 Does the effectiveness of the *Stress* treatment differ by session order?

Here, we check whether there are systematic differences in the treatment responses between participants who were in the *Stress* condition in their first session and those who were in this condition in their second session.

To this end, we conduct simple t-tests comparing the increases between baseline and treatment for heart rate and salivary cortisol. We also compare the average of self-reported stress post-treatment between the two groups. The so-called 'responder' rate, which is the threshold of a minimum of 2.5nmol/l increase in cortisol between baseline and treatment, is 66.67% for those who received the *Stress* treatment in their first session, and 53.42% for those who received it in their second session. However, this difference is not statistically significant (two-sided t-test, p-value= 0.109). For the heart-rate increase, the difference is not statistically significant either (two-sided t-test, p-value= 0.893), being on average 22.11 beats/min for *Stress* in participants' first session, and 22.48 beats/min on average for *Stress* in participants' second session. Finally, the statistical difference in the post-treatment self-reported stress is also not significant (two-sided t-test, p-value = 0.632), being on average 3.74 points for *Stress* in participants' first session, and 3.81 points on average for *Stress* in participants' second session.

3.D.2 Alternative clustering of the standard errors

Table 3.D.1. Within-subject comparison in risk-taking (alternative clustering of the standard errors)

	Change in Standardized RP										
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)	(5) (Gain)	(6) (Loss)	(7) (Full)	(8) (Full)	(9) (Gain)	(10) (Loss)	(11) (Full)
Constant	0.001	-0.002	0.003	0.001	-0.002	0.003	0.002	0.001	-0.002	0.003	0.002
	(0.005)	(0.006)	(0.006)	(0.004)	(0.008)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)
Observations	3405	1721	1684	3405	1721	1684	142	3405	1721	1684	142
Cluster-Level	Individual	Individual	Individual	Lottery	Lottery	Lottery	Lottery	–	–	–	–
Observation-Level	Decision	Decision	Decision	Decision	Decision	Decision	Participant	Decision	Decision	Decision	Participant
R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The analysis is at the participant-decision level in Columns (1) - (6), (8) - (10), and at the participant-level in Columns (7) and (11). Columns (1), (4), (7), (8), (11) report estimates for the full sample, Columns (2), (5), (9) report estimates for the gain domain, and Columns (3), (6), (10) for the loss domain. The values in parentheses represent standard errors clustered at the individual level (142 clusters for Full and Loss sample, 141 for Gain sample) for Columns (1) - (3), at the lottery level (28 clusters, 14 clusters for samples by domain) for Columns (4) - (7), and no clustering for Columns (8) - (11). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.3 Analysis as a fixed effects regression

Table 3.D.2 replicates the results presented in Table 3.3, i.e., implementing a fixed effects regression, replicating the within-individual changes analysis. Instead of the standard fixed-effects XTREG command in STATA, we use the high-dimensional fixed effects regression, implemented in the command REGHDFE (Correia, 2014, 2016), as it allows for clustering standard errors. This specification absorbs individual fixed effects, and subsumes lottery fixed effects by design. Standard errors are clustered at the group level. The table is organized in the same way as Table 3.3 in the main analysis, with Columns (1)-(3) pertaining to the lottery level, and Column (4) to the individual level.

Table 3.D.2. Within-subject comparison in risk-taking (fixed effects specification)

	Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Stress	0.001 (0.005)	-0.002 (0.006)	0.003 (0.006)	0.002 (0.005)
Constant	0.014*** (0.002)	-0.026*** (0.003)	0.054*** (0.003)	0.011*** (0.002)
Observations	6810	3442	3368	284
Observation-Level	Decision	Decision	Decision	Participant
R − within	0.000	0.000	0.000	0.001

Notes: All columns report estimates from individual Fixed Effects (FE) regressions. The dependent variable is the standardized risk premium (RP, see Section 3.2.3). The analysis is at the participant-decision level in Columns (1) - (3), and at the participant-level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). All specifications include participant fixed effects. The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

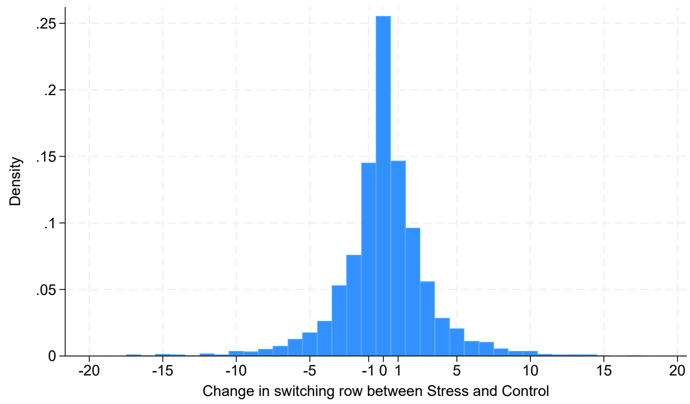
3.D.4 Order effects

Table 3.D.3. Within-subject comparison in risk-taking by order of the *Stress* treatment

	Change in Standardized RP			
	(1) (ST-CO)	(2) (CO-ST)	(3) (ST-CO)	(4) (CO-ST)
Constant	0.004 (0.007)	-0.002 (0.007)	0.004 (0.007)	-0.000 (0.007)
Observations	1687	1718	69	73
Observation-Level	Decision	Decision	Participant	Participant
Treatment Order	Stress-Control	Control-Stress	Stress-Control	Control-Stress
R^2	0.000	0.000	0.006	0.000

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The analysis is at the participant-decision level in Columns (1) and (2), and at the participant-level in Columns (3) and (4). Columns (1) and (3) report estimates for participants who were in the *Stress* condition first, and Columns (2) and (4) for participants who were in the *Control* condition first. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.5 Change in switching row



Notes: This figure plots the difference between the switching row in the *Stress* and *Control* conditions for all participants and lotteries. For 54.7% of the 3,405 observed participant-lottery combinations, the row in which the participant switched preference between the safe payment and lottery in the *Stress* condition was at most one row apart from their respective choice in the *Control* condition.

Figure 3.D.1. Difference between switching row in *Stress* and *Control*

3.D.6 Analyses assuming a functional form for risk preferences

3.D.6.1 Main model

We write the equivalence relation between the safe payment and lottery G as:

$$U(CE_G) = U(x_L)w(p_L) + U(x_l)(1 - w(p_L)), \tag{3.D.1}$$

where x_L , p_L , and x_l, indicate the high outcome, its probability, and the low outcome. In order to estimate $U(\cdot)$ and $w(\cdot)$, we follow Bruhin, Fehr-Duda, and Epper (2010) and Murad, Sefton, and Starmer (2016) and assume a sign-dependent CRRA utility function,

$$U(x) = \begin{cases} x^{\alpha}, & \text{if } x \geq 0, \\ -(-x)^{\beta}, & \text{otherwise,} \end{cases}$$

and the linear-in-log-odds probability weighting function proposed by Goldstein and Einhorn (1987) and Lattimore, Baker, and Witte (1992),

$$w(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}}$$

$$\delta \geq 0, \gamma \geq 0, \gamma < 1.$$

The advantage of this specification is that the two parameters have a clear interpretation: the δ parameter captures the elevation of the probability weighting function, while γ captures its curvature.

3.D.6.2 Estimation of the above model on our data

Table 3.D.4. Structural estimation of parameters

	(1)
alpha	0.93***
	(0.10)
delta	0.90***
	(0.05)
gamma	0.47***
	(0.01)
beta	1.05***
	(0.11)
Observations	6,810

Notes: Least-squares estimation of parameters based on 6,810 observations by 142 individuals, using choices made in the *Stress* and in the *Control* condition. Starting values follow Bruhin, Fehr-Duda, and Epper (2010) for the pooled EUT parameters in the gain domain: $\alpha = 0.981$, $\delta = 0.911$, $\gamma = 0.943$, and for losses $\beta = 1.015$. We assume that parameters are homogeneous across participants, and that δ and γ do not differ across domains. γ is constrained to lie in $(0,1)$, while α, β, and δ are restricted to positive values. The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

We use the estimated parameters reported in Table 3.D.4 to compute the predicted certainty equivalent for each decision in the data, treating the estimates as the true underlying structural parameters. For each lottery, this yields a model-implied certainty equivalent that represents the choice participants would make if they behaved exactly according to Equation (3.D.1).

We then compare these predicted certainty equivalents to participants' actual choices. For each observation, we compute the deviation from the predicted value and square this deviation for comparability across positive and negative residuals. The mean squared deviation is 28.9 (SD 141) in the *Control* condition and 29.1 (SD 164) in the *Stress* condition, a difference that is not statistically significant. Within domains, the pattern is similar. In the loss domain, the squared deviation is 32.2 in the *Control* and 30.2 in the *Stress* condition; in the gain domain, it is 25.7 in the *Control* and 28.0 in the *Stress* condition. None of these differences are statistically significant.

3.D.6.3 Estimation of the above model on our data, separately by Stress and Control

We also estimate the structural model separately using only data from the *Stress* and only from the *Control* condition. Estimated parameters change little, and the resulting model fit improves only minimally. Using condition-specific parameters yields average squared deviations of 28.9 (SD 140) in the *Control* and 29.1 (SD 166) in the *Stress* condition.

Table 3.D.5. Structural estimation of parameters, separately by condition

	(1) Control	(2) Stress
alpha	0.89***	0.97***
	(0.14)	(0.14)
delta	0.93***	0.86***
	(0.07)	(0.07)
gamma	0.47***	0.47***
	(0.01)	(0.01)
beta	1.09***	1.02***
	(0.15)	(0.15)
Observations	3,405	3,405

Notes: Least-squares estimation based on the subsamples of choices made in the *Control* (Column 1) and *Stress* (Column 2) condition. Starting values follow Bruhin, Fehr-Duda, and Epper (2010). Parameter restrictions are identical to Table 3.D.4. The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.6.4 Estimation of deviations based on Bruhin, Fehr-Duda, and Epper (2010) estimated parameters

As an alternative exercise, we follow the approach of Bruhin, Fehr-Duda, and Epper (2010) and classify individuals according to whether their observed certainty equivalents are better explained by an Expected Utility (EUT) model or by Cumulative Prospect Theory (CPT). For this purpose, we treat the pooled structural parameters reported in Bruhin, Fehr-Duda, and Epper (2010) as fixed type-specific preference parameters. For EUT types, we use the pooled gain-domain parameters $\alpha_g^{EUT} = 0.981$, $\gamma_g^{EUT} = 0.943$, $\delta_g^{EUT} = 0.911$ and the loss-domain parameters $\beta_l^{EUT} = 1.015$, $\gamma_l^{EUT} = 0.950$, $\delta_l^{EUT} = 1.072$. For CPT types, we use the pooled gain parameters $\alpha_g^{CPT} = 0.941$, $\gamma_g^{CPT} = 0.377$, $\delta_g^{CPT} = 0.926$ and loss parameters $\beta_l^{CPT} = 1.139$, $\gamma_l^{CPT} = 0.397$, $\delta_l^{CPT} = 0.991$. As in the original study, curvature and probability-weighting parameters are allowed to differ by domain, and given a type, parameters are assumed to be homogeneous across individuals.

For each participant, we compute the model-implied certainty equivalent for every lottery using both the EUT parameter vector and the CPT parameter vector, and we classify individuals according to which model yields the lower sum of squared deviations from their actual choices. Using this classification rule, 88 participants are classified as CPT types and 54 as EUT types, which is consistent with the strong curvature in probability weighting estimated in our own structural model above.

Using each participant's assigned type, we compute model-predicted certainty equivalents and compare them to their actual choices. The resulting mean squared deviation is similar across treatments, with 25.2 (SD 121.3) in the *Control* and 25.7 (SD 141.0) in the *Stress* condition.

Finally, we repeat this exercise separately for data from the *Control* and *Stress* conditions. Using only the *Control*-condition choices, 85 individuals are classified as CPT and 57 as EUT, with a resulting mean squared deviation of 23.9 (SD 120.4). Using only the *Stress* condition data, 83 individuals are classified as CPT and 59 as EUT, with a mean squared deviation of 24.6 (SD 140.4). These results suggest a modest but only minor improvement in fit when type classification is performed separately by condition.

3.D.7 Chronic stress and acute stress interaction

Next, we examine whether participants' chronic stress levels moderate the observed relationship between acute stress and risk-taking behavior. Previous research has shown that chronic stress, defined as stress sustained over extended periods, influences how individuals respond to acute stressors (Chida and Hamer, 2008).[57] Hence, we a priori expect that chronic stress may influence how acute stress affects risky choice.[58]

57. Ockenfels et al. (1995) find no correlation between chronic stress and an individual's reaction to an acute stressor, whereas Matthews, Gump, and Owens (2001) and Kudielka et al. (2006) find that individuals with higher chronic stress levels tend to show attenuated physiological responses to acute stressors.

58. To our knowledge, no prior study has directly examined whether chronic stress moderates the effect of acute stress on risk-taking. Some studies have analyzed the direct effect of chronic stress on risk-taking, with mixed results (e.g., Kandasamy et al., 2014; Ceccato, Kudielka, and Schwieren, 2016).

We collected two validated markers of chronic stress: (i) hair cortisol concentration (HCC), a biological indicator of cumulative cortisol exposure over recent months (Stalder and Kirschbaum, 2012), and (ii) the Trier Inventory of Chronic Stress (TICS) questionnaire (Schulz and Schlotz, 1999).

A total of 71 participants provided hair samples. Applying our preregistration exclusion criteria and using our main analysis sample, we retained 60 participants for analysis. Following standard practice, we log-transform HCC values to mitigate the influence of outliers (Dettenborn et al., 2012). The average log-transformed HCC is 1.76 (SD = 0.72; min = 0.06; max = 3.67) in the first segment, and 1.67 (SD = 0.85; min = -.60; max = 3.41) in the second segment. These values are comparable to previous studies (Staufenbiel et al., 2013) and fall within a range typically observed in healthy populations, indicating that our sample does not include individuals with clinically elevated chronic stress.

The TICS questionnaire is available for all 142 participants. It has 39 items rated on a five-point agreement scale and comprises the following sub-scales: 'work overload' (8 items), 'worries' (6 items), 'social strain' (6 items), 'lack of social recognition' (8 items), 'work discontent' (5 items), and 'intrusive memories' (6 items). We reverse-code items for which agreement indicates the absence of chronic stress so that higher values uniformly reflect greater chronic stress across all items. We compute indices for all six sub-scales: 'work overload' (mean = 2.95, SD = 0.79; min = 1.5; max = 4.75), 'worries' (mean = 2.91, SD = 0.83; min = 1.17; max = 4.83), 'social strain' (mean = 2.20, SD = 0.59; min = 1; max = 4.33), 'lack of social recognition' (mean = 1.56, SD = 0.49; min = 0.5; max = 3.5), 'work discontent' (mean = 2.33, SD = 0.76; min = 0.8; max = 4.4), and 'intrusive memories' (mean = 2.73, SD = 0.88; min = 1; max = 5). We also compute a summary index by averaging responses across all 39 items (mean = 2.43, SD = 0.53; min = 1.18; max = 3.82).

First, we assess the relationship between chronic stress and acute stress responses by regressing the increase in cortisol between the first sample after onset of the treatment phase (Sample 3) and baseline (Sample 2) (compare Appendix 3.C.2.2), on each of our chronic stress markers. Table 3.D.6 presents the results. We find no significant correlations between participants' salivary cortisol responses to treatment and any of our chronic stress markers. This suggests that in our sample, there is no evidence for chronic stress predicting acute cortisol responses.

Table 3.D.6. Correlation between acute and chronic stress

	Increase in Salivary Cortisol								
	(1) (Hair)	(2) (Hair)	(3) (Full)	(4) (Full)	(5) (Full)	(6) (Full)	(7) (Full)	(8) (Full)	(9) (Full)
log(HCC1)	-0.508 (1.063)								
log(HCC2)		-0.708 (0.828)							
Total TICS			-0.947 (1.066)						
Work Overload				-0.642 (0.847)					
Worries					-0.167 (0.670)				
Social Strain						-2.003 (1.203)			
Lack of Soc./ Recognition							0.238 (1.299)		
Work Discontent								0.433 (0.739)	
Intrusive Memories									-0.826 (0.615)
Constant	6.743*** (1.905)	7.030*** (1.579)	7.968*** (2.563)	7.560*** (2.547)	6.150*** (1.868)	10.063*** (2.914)	5.293*** (1.848)	4.654** (1.657)	7.924*** (1.754)
Observations	60	60	142	142	142	142	142	142	142
R^2	0.003	0.009	0.006	0.007	0.000	0.036	0.000	0.003	0.013

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *difference* between salivary cortisol (in nmol/l) in the first sample after onset of the treatment, Sample 3, and the baseline sample, Sample 2 (compare Appendix 3.C.2.2). Columns (1) and (2) include only data from participants who provided a hair sample, while Columns (3) - (9) include the full sample. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

We preregistered a robustness check, in which we account for participants' chronic stress by interacting the *Stress* condition with a chronic stress measure, HCC_i, $TICS_i$, or the respective TICS sub-scales. Note that in our setup, the coefficient estimate of the chronic stress measure is equivalent to the coefficient of an interaction term in a classical moderation analysis. Table 3.D.7 presents the corresponding results. Note that we focus on the first hair segment (i.e., HCC1) due to its high correlation with HCC2 (rho = 0.914).

With the exception of the 'social strain' sub-scale (Column (5)), all results are in line with our main results, which cannot reject that there is no effect of stress on risk-taking. Results in Column (5) suggest some heterogeneity in treatment effects depending on self-reported experience of social strain within the past year. After correcting the coefficients relating to the six TICS sub-scales for multiple hypothesis testing (the respective p-values are provided in Table 3.D.7), we find an effect significant at the 10% level. We are cautious about giving this result too much weight, as our overall evidence overwhelmingly supports a nil effect.

Table 3.D.7. Within-subject change in standardized RP and chronic stress

	Change in Standardized RP							
	(1) (Hair)	(2) (Hair)	(3) (Full)	(4) (Full)	(5) (Full)	(6) (Full)	(7) (Full)	(8) (Full)
Constant	0.025 (0.028)	-0.023 (0.022)	-0.006 (0.019)	-0.011 (0.017)	-0.035* (0.017)	-0.011 (0.016)	-0.004 (0.016)	-0.013 (0.012)
log(HCC1)	-0.011 (0.014)							
Total TICS		0.010 (0.009)						
Work Overload			0.002 (0.006) [0.643]					
Worries				0.004 (0.006) [0.848]				
Social Strain					0.016** (0.008) [0.079]			
Lack of Soc./ Recognition						0.008 (0.010) [0.988]		
Work Discontent							0.002 (0.007) [0.919]	
Intrusive Memories								0.005 (0.004) [0.898]
Observations	1423	3405	3405	3405	3405	3405	3405	3405
Observation-Level	Decision	Decision	Decision	Decision	Decision	Decision	Decision	Decision
R^2	0.002	0.001	0.000	0.000	0.003	0.000	0.000	0.001

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The values in round parentheses represent standard errors clustered at the participants' group level (23 clusters). The values in square parentheses represent Romano-Wolf p-values adjusted for multiple hypothesis testing. Columns (1) and (2) include only data from participants who provided a hair sample, while Columns (3) - (9) include the full sample. The statistical significance levels (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$) are based on the clustered, non-adjusted standard errors.

3.D.8 Analysis of decision timeouts

Here, we analyze the extent to which the timeout of decisions affects our results. First, we check whether decisions are more likely to time out under certain conditions. Second, we present our main analyses limited to individuals who did not have any timed-out decisions.

In Table 3.D.8, we present the probit estimates from regressing whether a decision was timed out on a lottery's skewness (dummy for respective skewness vs. all the other categories), stake sizes (dummy for respective stake size vs. all the other categories), an indicator for first decision in session, participants' session (dummy for whether it is a participant's first session), female (dummy), and a dummy for the *Stress* condition.

According to this table, we see that 'easier' MPLs (skewness 50:50) and low stakes (10 vs. 20) decrease the probability of a timeout, likely because they are easier to process for individuals. Accordingly, we find that high-stakes lotteries (50 vs. 150) increase the probability of a timeout. Furthermore, lotteries in the beginning, be it the first decision or participants' first session, are more likely to time out. Also, female participants are less likely to have a timed-out decision. However, most importantly, being in the *Stress* condition does not increase the probability of a timeout.

Table 3.D.8. Probability of a timeout

	Probability of a Timeout					
	(1)	(2)	(3)	(4)	(5)	(6)
Skewness 5/95	0.164**					
	(0.065)					
Skewness 10/90	0.106					
	(0.081)					
Skewness 25/75	0.217***					
	(0.067)					
Stakes 10/40		0.042				
		(0.064)				
Stakes 20/50		0.114*				
		(0.063)				
Stakes 50/150		0.234**				
		(0.092)				
First Decision			0.392***			
			(0.100)			
First Session				0.606***		
				(0.050)		
Female					-0.301***	
					(0.046)	
Stress						-0.062
						(0.045)
Constant	-1.700***	-1.635***	-1.577***	-1.932***	-1.432***	-1.529***
	(0.053)	(0.051)	(0.023)	(0.041)	(0.029)	(0.031)
Observations	7952	7952	7952	7952	7952	7952
Observation-Level	Decision	Decision	Decision	Decision	Decision	Decision
$Pseudo - R^2$	0.003	0.002	0.004	0.045	0.012	0.001

Notes: All columns report estimates from probit regressions. The dependent variable is a dummy variable that takes the value of one if a decision timed out. The values in parentheses represent standard errors. The omitted category in Column (1) is an indicator for a lottery with a skewness 50/50. The omitted category in Column (2) is an indicator for a lottery with stakes of 10/20. The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

Subsequently, we limit our analysis to participants who did not have a timeout in any of the 56 choices, which applies to 42 participants. The results are in Table 3.D.9. We find that our main results are qualitatively confirmed.

Table 3.D.9. Main analyses limited to individuals who have submitted all 56 choices

	Change in Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Constant	0.010	0.000	0.020*	0.010
	(0.010)	(0.013)	(0.012)	(0.010)
Observations	1176	588	588	42
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.003	0.000	0.013	0.037

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. Column (1) reports estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.9 Further preregistred analyses

In this Section, we provide the results from further preregistered analyses. An overview of our deviations from the preregistration is provided in Appendix 3.E.

3.D.9.1 Further preregistered tests of the between-subject analysis

First, we preregistered a t-test and a Wilcoxon-Mann-Whitney test to compare the risk-taking behavior between participants who were first assigned to the *Stress* condition and participants who were first assigned to the *Control* condition. According to a two-sided t-test, there is a statistically significant difference between the means (two-sided t-test, p-value < 0.001); however, it is marginally not significant according to a Wilcoxon-Mann-Whitney test (p-value $= 0.006$). These results are broadly in line with our between-subject analysis (see Section 3.3.2.4).

3.D.9.2 Inclusion of individuals who violated the behavioral rules or self-reported cortisol-effective health conditions

For our main analysis, we exclude individuals who either self-report a violation of behavioral rules or health conditions, denoted by the dummy VBH_i, that may compromise cortisol measures. We preregistered a robustness check, in which we keep these individuals by interacting the *Stress* condition with VBH_i. Note that in our setup, the coefficient estimate of the VBH_i dummy is equivalent to the coefficient of an interaction term in a classical moderation analysis. Table 3.D.10 presents the corresponding results. This analysis comprises data from 155 individuals and 3,674 choice pairs.

Table 3.D.10. Within-subject change in standardized RP (including VBH)

	Change in Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Constant	0.001	-0.002	0.003	0.002
	(0.005)	(0.006)	(0.006)	(0.005)
VBH	-0.013	-0.032	0.001	-0.009
	(0.024)	(0.032)	(0.022)	(0.025)
Observations	3674	1843	1831	155
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.000	0.002	0.000	0.002

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The analysis is at the participant-decision level in Columns (1) - (3), and at the participant-level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.9.3 Inclusion of stochastically dominated decisions

We preregistered a robustness check incorporating stochastically dominated choices. If we assume that some of the difference between risk-taking behavior under *Stress* and *Control* is driven by a different rate of error taking, keeping stochastically dominated decisions is reasonable. In this case, we have 3,554 choice pairs.[59]

Table 3.D.11 presents the same analysis as in Table 3.3, i.e., according to Equation (3.1) but on a sample including stochastically dominated choices.

59. In comparison to the main sample (compare Table 3.1), the 3,554 observations are comprised as follows: 7,952 potential lottery choices (7,952=142 participants×28 lotteries×two sessions) minus 473 timed out decisions, minus 371 choices, for which we do not have a corresponding choice in the other condition (because of it being excluded due to a timeout).

Table 3.D.11. Within-subject change in standardized RP (incl. dominated decisions)

	Change in Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Constant	0.001	-0.006	0.009	0.001
	(0.005)	(0.006)	(0.006)	(0.005)
Observations	3554	1806	1748	142
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.000	0.001	0.002	0.000

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the *change* in the standardized risk premium (RP, see Section 3.2.3), between the *Stress* and *Control* condition. The analysis is at the participant-decision level in Columns (1) - (3), and at the participant-level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.9.4 Decision order effects

We preregistered to control our main specifications for order effects. Including order effects assumes that decision behavior differs between the first and last MPLs shown. Hence, not all variation in decision behavior would be captured by lottery fixed effects. We, therefore, present a version of Table 3.5 including a factor variable for the decision order from 1 to 28.

Table 3.D.12. Between-subject comparison in risk-taking (participants' first session)

	Standardized RP					
	(1) (Full)	(2) (Full)	(3) (Full)	(4) (Full)	(5) (Gain)	(6) (Loss)
Stress	-0.029**	-0.029**	-0.028**	-0.031**	-0.038***	-0.024
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.017)
Female			0.014	0.017	0.007	0.027
			(0.012)	(0.013)	(0.013)	(0.016)
Age			0.000	0.000	-0.001	0.002
			(0.002)	(0.001)	(0.001)	(0.002)
Constant	0.030***	0.012	0.003	-0.150***	-0.090**	-0.160***
	(0.007)	(0.024)	(0.040)	(0.037)	(0.041)	(0.057)
Lottery FE				X	X	X
Dec. Number FE		X	X	X	X	X
Observations	3405	3405	3405	3405	1721	1684
R^2	0.004	0.011	0.012	0.452	0.450	0.434

Notes: All columns report estimates from ordinary least squares (OLS) regressions. The dependent variable is the standardized risk premium (RP, see Section 3.2.3). Columns (1) - (4) report estimates for the full sample, Column (5) reports estimates for the gain domain, and Column (6) for the loss domain. The specifications in Columns (2) - (6) control for a female dummy and participants' age, Columns (2) - (6) include decision order effects, and Columns (3) - (6) additionally include lottery fixed effects. The values in parentheses represent standard errors clustered at the participants' group level (23 clusters). The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

3.D.9.5 Instrumental variable approach

Strictly speaking, in our main analysis, we estimate an intention-to-treat (ITT) effect since not all participants exposed to the *Stress* condition are actually physiologically stressed.

We, therefore, check the robustness of our main result by estimating a causal average treatment effect (ATT), (compare, e.g., Cahlíková and Cingl, 2017). To this end, we run a two-stage instrumental variable regression, where we instrument physiological stress with experience of the *Stress* condition. As the measure of physiological stress, we use participants' increase in cortisol between the first sample after onset of the treatment phase (Sample 3) and baseline (Sample 2) (compare Appendix 3.C.2.2). Table 3.D.13 presents the results of the second stage regression, where we regress the risk premium on the instrumented physiological stress. For the first stage, we use the same regression for all specifications, regressing the salivary cortisol increase measured for a given participant in a given session on whether the session was in the *Stress* condition.

Table 3.D.13. Within-subject comparison in risk-taking (IV)

	Standardized RP			
	(1) (Full)	(2) (Gain)	(3) (Loss)	(4) (Full)
Increase in Salivary Cortisol	0.000	-0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	0.130***	-0.006***	0.265***	0.129***
	(0.001)	(0.001)	(0.001)	(0.002)
Observations	6810	3442	3368	284
Observation-Level	Decision	Decision	Decision	Participant
R^2	0.116	0.171	0.149	0.902
Stress (first stage)	6.353***			
	(0.541)			

Notes: All columns report the second-stage estimates from two-stage least squares (2SLS) regressions. The dependent variable is the standardized risk premium (RP, see Section 3.2.3). The analysis is at the participant-decision level in Columns (1) - (3), and at the participant-level in Column (4). Columns (1) and (4) report estimates for the full sample, Column (2) reports estimates for the gain domain, and Column (3) for the loss domain. The first stage is common across all four specifications and regresses the individual increase in cortisol between the first sample after onset of the treatment phase (Sample 3) and baseline (Sample 2) on the *Stress* condition dummy and individual fixed effects (N=284). The values in parentheses represent standard errors. The statistical significance is reported on the * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ level.

Appendix 3.E Research transparency

This study was preregistered on AsPredicted.org under #100204, https://aspredicted.org/t3pv74.pdf. The preregistration includes the main hypotheses, the experimental conditions, measurement of the key dependent variables, the planned sample size, the main analyses, and the supporting analyses. In the following, we document where we deviate from the preregistration.

- Sample Size
  - We had preregistered a target of 150 participants, whom we planned to group in 20 experimental groups attending two sessions each (i.e., 40 sessions in total). Due to lower turnout than expected (i.e., fewer than 8 participants per session), we recruited and conducted sessions with an additional three groups, i.e., 23 groups and 46 total sessions. The sample size of 142 participants (in the main sample) differs slightly from the preregistered sample size due to the preregistered exclusion criteria, which were only applied after the data collection.
- Analysis
  - We had preregistered that we would control for gender and age in all analyses. In analyses, which refer to within-individual changes (e.g., Table 3.3), all individual-level characteristics are by design controlled for, and the additional inclusion of gender and age controls is not possible and nonsensical. Our fixed effects analysis in Appendix 3.D.3 includes individual fixed effects, which also makes the inclusion of gender and age controls nonsensical.
  - We preregistered our main analysis to rest on $\overline{RP} = (EV - CE)/|EV|$ as the main dependent variable and to use an individual × lottery fixed effects regression for analysis. We resort to $\Delta\overline{RP}_{il}$, which is the difference between the $\overline{RP}_{il}$ in *Stress* and *Control*, as our main dependent variable. Furthermore, we regress our main dependent variable on a constant. We use this approach for ease of exposition, but it yields the same coefficient estimates as the preregistered approach (compare Section 3.3.2.2 and the results of the fixed effects analysis in Appendix 3.D.3).
  - We did not preregister any clustering of the standard errors, but we clustered them in our main analysis. However, we provide non-clustered standard errors in Appendix 3.D.2.
  - In Appendix 3.D, we provide a range of results that were not preregistered, for instance, in Appendix 3.D.5. These additional results aim to provide further insight into the data.
  - We preregistered several secondary analyses (last paragraph in the preregistration) in addition to those presented in the Appendix, including further controls and heterogeneity tests. Given that we preregistered these analyses

as secondary and they do not provide meaningful insights into our results, we do not include them for expositional reasons. These results are available from the authors upon request.

Appendix 3.F Screenshots of the experiment

The experiment was conducted in German and later translated into English. The screenshots of the experiment as well as the translations can be found in the digital appendix on the bonndoc publication server:
https://nbn-resolving.org/urn:nbn:de:hbz:5-91417

Note that the screenshots label screens shown to participants in their first session as 'Week 1', since this was, for participants, the session they attended in the first week of their experiment experience. Conversely, those shown to participants in their second session are referred to as 'Week 2', since this was, for them, the second week of their experiment experience. The 'Week 1' and 'Week 2' labels do not refer to the calendar weeks in which sessions took place, i.e., we often hosted several 'Week 1' and 'Week 2' sessions within the same calendar week, and the period over which experimental sessions took place was longer than two weeks.