EXPERIMENTS ON GENDER AND SELECTION

MONIKA POMPEO

Thesis submitted to the University of Nottingham
for the Degree of Doctor of Philosophy

Supervised by:
Matthias Dahm, Daniele Nosenzo & Alex Possajennikov

School of Economics
University of Nottingham
November 2021

Abstract

Chapter 1 provides a general introduction to the themes and methods used in the thesis.

Chapter 2 reports an experiment that examines the interaction between performance payments and selection into the mission sector. Subjects are asked to choose between a high-powered profit sector and a mission sector, where I vary the type of incentive depending on the treatment. I find that, depending on the setting, performance-based payments either crowd out intrinsic motivation or are able to attract more productive and altruistic individuals to the mission sector. This difference depends on how the options are structured and the type of incentives offered in each sector.

Chapter 3 reports an experiment studying the choice to make others compete. Contrary to our initial expectations, we find no evidence of discrimination when it comes to the choice to make others compete. This effect does not depend on the gender of the principal making the tournament entry decision, but it is associated with his or her own risk and competitive preference. We also find mixed evidence when it comes to the efficiency loss when subjects are made to enter competition instead of choosing it themselves. In particular, the choices made by principals lead to lower payoff maximising decisions and lower performance. However, the latter is not robust to controlling for subjects' ability in the real-effort task.

Chapter 4 studies the effect of gender norms as a coordination device in the context of a battle of the sexes game. In particular, the online experiment was run in India and Norway, two countries on the opposite side of the spectrum when it comes to gender equality. We find that subjects are more aggressive (they are more likely to choose the strategy that gives them more points) towards females in India, but in particular, low-education females are more aggressive towards other females. In Norway, we find that highly educated males tend to be more aggressive towards other males. We argue that the driver behind the results is that subjects experience different levels of inequality aversion depending on the gender of the opponent, which interacts with gender norms between countries.

Chapter 5 summarises the findings and concludes the thesis.

Acknowledgements

My sincere gratitude goes to my supervisors, Matthias Dahm, Daniele Nosenzo and Alex Possajennikov, for their support and guidance throughout the PhD journey. I am particularly thankful for their patience and ability to read through my scrambled words and transform them into sound arguments. They made my supervision meetings a great learning environment, filled with discussions and, importantly, laughter.

My gratitude also extends to the rest of the Centre for Decision Research and Experimental Economics (CeDEx) for their valuable suggestions and for making me feel that I could always knock on their door to talk about research. A special acknowledgement goes to Suzanne for lending me her shoulder to cry from time to time and to Jose for teaching me how to program while imparting life lessons.

Finally, I would like to thank my partner Gian Luca for all the support he has given me through the years.

Contents

Abstract i

Acknowledgments ii

List of Figures v

List of Tables vi

Chapter I - Introduction

Chapter II - The Sorting Effect of High-Powered Incentives in the Mission-Oriented Sector

1 Introduction . . . 8
2 Related Literature . . . 11
3 Experimental Design . . . 14
3.1 Online Stage . . . 15
3.2 Lab Experiment . . . 16
3.3 Procedures . . . 18
4 Results . . . 18
4.1 Behavioural Measures . . . 18
4.2 Selection into the mission-oriented sector . . . 20
4.3 Output . . . 28
4.3.1 Charitable donations . . . 31
5 Discussion and conclusions . . . 33
A Supplementary Tables and Figures . . . 36
A.1 Tables . . . 36
A.2 Figures . . . 37

Chapter III - Choosing Competition on Behalf of Someone Else

1 Introduction . . . 51
2 Related literature . . . 54
2.1 Willingness to compete . . . 54
2.2 Taking decisions for others . . . 55
2.3 Efficiency . . . 57
3 Experimental design and procedures . . . 58

3.1 Agents/Decision-makers . . . 60
3.2 Principals . . . 61
4 Results . . . 62
4.1 Summary statistics . . . 63
4.2 Gender gap in competitiveness . . . 64
4.3 Determinants of the principal's choice . . . 68
4.3.1 Principal's characteristics . . . 68
4.3.2 Principal's evaluation of the agent's relative performance . . . 70
4.4 Efficiency . . . 72
4.4.1 Ex-ante efficiency . . . 72
4.4.2 Ex-post efficiency . . . 74
5 Conclusions . . . 76
A Supplementary Tables and Figures . . . 79
A.1 Descriptive statistics and performance in the task . . . 79
A.2 Gender gap in competitiveness . . . 84
A.3 Determinants . . . 85
A.4 Robustness checks . . . 88

Chapter IV - Gender Effects in the Battle of the Sexes: a Tale of Two Countries

1 Introduction . . . 91
2 Experimental Design and Procedures . . . 95
2.1 Sample description . . . 98
3 Hypotheses . . . 99
4 Results . . . 103
4.1 Discussion . . . 109
5 Concluding Remarks . . . 113
A Supplementary Tables . . . 115
A.1 Differences in expectations according to the opponent's gender (female – male opponent) . . . 122
B Probit estimates . . . 123

Chapter V - Conclusion

References **132**
References . . . 133

Instructions **146**

List of Figures

Chapter II - The Sorting Effect of High-Powered Incentives in the Mission-Oriented Sector

1 Summary of the structure of the lab-based part of the experiment 15
2 Change in the composition of the mission sector between treatments 21
3 Likelihood of choosing each option based on productivity and altruism . . . 24
4 Likelihood of choosing the mission sector based on productivity and altruism 26
5 Average number of matrices solved in the first and second stage across treatments 29
6 Average ex ante and ex post charitable donations 32
A.1 Coefficient plot examining the determinants behind the subject's choice to inscribe to the second stage of the experiment 45
A.2 Distribution of standardised preference measures 46
A.3 Marginal effects examining the interaction between treatments and productivity 47
A.4 Cumulative density function of individual output in the working stage of the experiment by sector and treatment 48

Chapter III - Choosing Competition on Behalf of Someone Else

1 Tournament entry by gender and treatment ($N = 416$) 64
A.1 CDFs of risk preferences in the lottery by gender and treatment 80
A.2 CDFs of the Stage 1 and Stage 2 performances by gender and treatment . . 82
A.3 Tournament entry by gender and treatment (N = 416) 85

Chapter IV - Gender Effects in the Battle of the Sexes: a Tale of Two Countries

1 Difference in probability of playing hawkish when facing a female opponent compared to a male opponent. 106
A.1 Difference in the probability of saying that the opponent will play hawkish when facing a female opponent compared to a male opponent. 122

List of Tables

Chapter II - The Sorting Effect of High-Powered Incentives in the Mission-Oriented Sector

1 Determinants of the choice of sector and incentive type for the High-Choice treatment . . . 23
2 Determinants of the choice of the mission sector . . . 25
3 Determinants of output in the mission sector . . . 30
A.1 Summary statistics . . . 36
A.2 Correlation between independent variables . . . 38
A.3 Composition of the mission sector across treatments . . . 39
A.4 Distribution of males and females across treatments . . . 40
A.5 Determinants of the choice of sector and incentive type for the High-Choice treatment . . . 41
A.6 Determinants of the choice of the mission sector for the High-Low treatment 42
A.7 Determinants of the choice of the mission sector for the High-Choice treatment 43
A.8 Determinants of the choice of the mission sector for the High-High treatment 44

Chapter III - Choosing Competition on Behalf of Someone Else

1 Main features of the experimental design . . . 59
2 Number of participants by treatment and role . . . 60
3 Tournament entry by treatment . . . 65
4 Principals' beliefs about agents' relative performance in Stage 2 in *GenderInfo* 71
5 Number of decision-makers and principals who maximise and do not maximise expected earnings, by treatment . . . 73
6 Performance profiles of subjects competing in Stage 3, by treatment . . . 73
7 Ex-post efficiency . . . 75
8 Stage 3 performance of those competing and winning, by treatment . . . 76
A.1 Descriptive statistics of subjects' characteristics by treatment . . . 79
A.2 Number of principals choosing piece-rate and tournament for their agent, by the principal's gender and treatment . . . 81
A.3 Tournament entry between treatments . . . 84
A.4 The influence of principal characteristics and preferences on the principal's tournament choice in *GenderInfo* . . . 86

A.5 The influence of principal characteristics and preferences on the principal's tournament in *NoGenderInfo* 87
A.6 Stage 3 performance of competing and winning agents when the principal's choice coincides with the agent's stated choice 88

Chapter IV - Gender Effects in the Battle of the Sexes: a Tale of Two Countries

1 Summary statistics 99
2 Description of coefficients 104
3 Propensity to choose the hawkish action- OLS estimates for India 105
4 Propensity to choose the hawkishish action - OLS estimates for Norway . . 108
5 Gender differences in the propensity to choose the hawkish action in the Gender Unknown treatment (men-women) 111
6 Propensity to choose the hawkish action– OLS estimates conditional on expected opponent's choice 112
A.1 Number of participants by country and treatment 115
A.2 Recall rates and study objective 115
A.3 Propensity to choose hawkish - pooled sample 116
A.4 Propensity to choose hawkish - Indian sample 117
A.5 Propensity to choose hawkish - Norwegian sample 118
A.6 Role of conservative/liberal gender attitudes – India 119
A.7 Role of conservative/liberal gender attitudes – Norway 120
A.8 Regression output for Figure 1 121
B.1 Probit estimates of the baseline model – Pooled sample 123
B.2 Probit estimates of the baseline model – India 124
B.3 Probit estimates of the baseline model – Norway 125
B.4 Role of conservative/liberal gender attitudes – India 126
B.5 Role of conservative/liberal gender attitudes – Norway 127
B.6 Propensity to choose hawkish – Probit estimates conditional on expected opponent's choice 128

CHAPTER I

Introduction

This dissertation presents three research projects. All projects fall within the fields of behavioural and experimental economics. The first combines an online and lab experiment, the second is a lab experiment, and the third is an online experiment. The research questions vary between them, but the two main themes of the dissertation are the study of selection and the effect of gender on different types of decisions. The project presented in Chapter 2 focuses on selection. The project presented in Chapter 3 focuses on gender and selection into competition. Finally, the project in Chapter 4 focuses on gender norms and how they vary between countries.

Chapter 2 is titled 'The Sorting Effect of High-Powered Incentives in the Mission-Oriented Sector'; it combines an initial online stage with a lab experiment. The chapter examines the effect of high-powered (or performance-based) incentives on selection into the mission sector. High-powered incentives have been often proposed as a way to increase productivity in the mission/public sector. Nonetheless, the use of such incentives is met with several challenges since output in this context is often hardly measurable. Also, they might lead to ignoring those aspects of the job that are not strictly related to the incentive and might affect intrinsic motivation by signalling that the relationship between the workers and the organisation is market-based (Burgess & Ratto, 2003). So far, the last aspect has been given very little attention. There exists an established link between productivity and the choice of performance-based payments (Dohmen et al., 2011; Eriksen, Kvaløy, & Luzuriaga, 2020), and between altruism and the choice of the mission sector (Serra, Serneels, & Barr, 2011; Lagarde & Blaauw, 2014). However, it is not clear a priori what will be the result of their interaction. Specifically, it is important to study whether it will affect the composition of the workforce. In this chapter, I focus on studying the effect of the use of high-powered incentives in the mission-oriented sector both in terms of selection and productivity.

I examine this issue with the use of an experimental design divided into two parts. The first is an online experiment where I elicit the subject's prosocial preferences using a series of survey measures. In the second, lab-based part of the experiment, the same subjects are invited to a seemingly unrelated lab experiment where they are asked to exert effort either for themselves (profit sector) or for a charity (mission sector). The design comprises three between-subjects treatments where I vary the type of incentive faced in the mission sector. In particular, I fix the type of incentive in the profit sector to be high-powered (performance-based) and have the subjects face either a low-powered incentive (fixed salary), a chosen power of incentive or a high-powered incentive in the mission sector.

When faced with a low-powered incentive in the mission sector and a high-powered incentive in the profit sector, the mission sector's selection process exhibits the standard problem; the fixed salary attracts less productive individuals who are more altruistic than

their profit-sector counterparts. When subjects can choose the power of the incentive in the mission sector, I find that those working under the high-powered incentive are more altruistic than those working in the profit sector, and more productive than those working in the mission sector under the low-powered incentive. When faced with a high-powered incentive in both sectors, I find that subjects choosing the mission sector tend to be just as productive but are not more altruistic than those in the profit sector. An important thing to note is that high-powered incentives do not necessarily lead to higher charitable donations. The fact that fewer subjects chose the mission-oriented sector under the high-powered incentive leads to lower contribution levels overall. My findings are particularly valuable given the recent rise in the number of programs aiming at increasing productivity in the mission/public sector (Duflo, Hanna, & Ryan, 2012; Binyaruka et al., 2015). They suggest that it is possible to screen for both altruism and productivity by using high-powered incentives, as long as the option to choose the low-powered one is present.

Chapter 3, titled 'Choosing competition on behalf of someone else', builds on (Niederle & Vesterlund, 2007) to examine the choice to make others compete. There is a vast, and growing , literature studying preferences for competition, but so far there are no papers looking at the decision to make others compete. The main research question is whether individual preferences for competition will extend to the case in which subjects choose competition on behalf of others. In other words, whether we would still observe a gender gap when individuals decide whether others should compete.

We assign subjects to three treatments varying whether they make the competitiveness decision themselves or whether it is made on their behalf by another randomly matched subject, referred to as the principal. We also vary the amount of information provided to the principal and, in particular, whether before making their decision they are informed, not only about the performance but also about their agent's gender. This last design feature allows us to causally examine the effect of gender when it comes to the decision to make others compete.

We find that principals are just as likely to send male and female agents into competition independently from whether the gender information is revealed. Also, we find that the gender of the principal does not matter. Nonetheless, individual preferences for competition and risk influence his or her decision to send the assigned agent into the tournament. Finally, we find mixed evidence when it comes to the efficiency loss when comparing those that chose for themselves and those that were selected to compete. Our findings suggest that principals make less payoff-maximising decisions and performance is lower in the treatments involving the role of the principal. However, the latter is not robust to controlling for past performance in the task. One of the potential explanations is that the emotional involvement of choosing

for someone else is lower than when choosing for oneself, which could lead to more objective tournament entry decisions. This also justifies creating roles within the organisation that could help people make this type of decision and, broadly speaking, extend the arguments favouring mentorship programs.

Chapter 4, titled 'Gender Effects in the Battle of the Sexes: a Tale of Two Countries', reports an online experiment run between two countries, Norway and India. The paper exploits a battle of the sexes game to study the use of gender as a coordination device and its interaction with gender norms in situations with a conflict of interest. In particular, the game is characterised by a context in which if the players want to avoid earning nothing, one of them has to concede and earn less than the other. In the paper, we refer to the hawkish strategy as giving more points and the dovish strategy as the one giving less to the player who chose it. The questions are which player will decide to concede, thus earning fewer points, and how do gender norms mediate this choice. To answer these questions, we employ a between-subjects design where players are randomly distributed between treatment and control conditions. In the control condition, they are matched with another player from the same country, but they are not informed of the gender. In the treatment conditions, they know both that it is a player from the same country and the gender. By examining the difference in behaviour between treatments and control groups, we can isolate the impact of the gender label on the subject's choice of strategy. The fact that we gathered information about the expectations about the other player's choice of strategy, also enables us to study potential drivers behind the player's choices, such as beliefs and preferences.

Our findings suggest that the country with the weakest level of gender equality, India, features more substantial gender effects. That is, Indian subjects are generally more likely to condition their behaviour on the gender of the opponent. However, when looking closer into the results, we notice that it is mainly driven by low-education females who are 20% more likely to choose the hawkish strategy when facing another woman. In Norway, we find that gender only matters for highly educated men, which are more likely to choose the aggressive strategy when facing other men. We examine potential drivers of behaviour starting by excluding that the effect is expectation-driven. We then proceed in the analysis to conclude that the effect is due to the interaction between inequality aversion and the gender of the opponent in a way that is mediated by culture and gender norms.

The broader implications of our findings suggest that interventions aiming at empowering women in countries with low gender equality should tackle the issue of gender entitlement first. If women perceive their lower status as something natural, policy should start by encouraging a stronger feeling of women's entitlement. For example, female role models have been found to impact aspiration and educational attainments for girls (Castilla, 2018; Beaman,

Duflo, Pande, & Topalova, 2012). Another potential venue could be engaging adolescents in classroom discussions about gender equality, which has been found to effectively shift gender attitudes towards more progressive stances in India (Dhar, Jain, & Jayachandran, 2018). Our results suggest that these types of interventions could be a step in the right direction to promote gender equality, especially in developing countries.

CHAPTER II

The Sorting Effect of High-Powered Incentives in the Mission-Oriented Sector

Abstract

This paper examines the effect of incentives on sorting across the mission-oriented and profit-oriented sectors and, more generally, studies the role played by individual characteristics in self-selection across sectors and incentive schemes. I employ an experimental design divided into an online stage and a lab experiment. In the online stage, I elicit subjects' risk aversion, discounting, trust, altruism, positive reciprocity, negative reciprocity, and self-control with the use of validated survey measures. Participants are then invited to the lab, where they are given the option to choose between working in the profit-oriented or the mission-oriented sector for different types of incentives. The incentive scheme associated with the profit sector is always high-powered (that is performance-based), whereas in the mission-oriented context, in one treatment it is low-powered (that is fixed), in another, subjects choose between a high- and a low-powered incentive, and in the last one, it is high-powered. I find that as long as high-powered incentives are used alongside low-powered ones, they have the potential to screen for more productive and altruistic individuals. However, when high-powered incentives are the only incentive type offered within the mission sector, they increase average productivity compared to the baseline but dilute prosocial motivation, thus weakening the link between altruism and the choice of the mission sector.

JEL Codes: C91, M52, J45

Keywords: sorting, mission-oriented sector, incentive schemes

1 Introduction

High-powered incentives (or performance-based payments) have often been proposed as a way to attract more productive workers in the mission-oriented sector.[1,2] The most recurring arguments against the use of this type of payment in the public sector have centred on the issues of multiple tasks, as the jobs often involve more than one aspect, which could lead subjects to ignore those that are not strictly related to the incentive; measurement, as output might not be easy to measure given that in most of the cases it is intangible; and intrinsic motivation, for high-powered incentives might signal that the relationship between the workers and the organisation is market-based (Burgess & Ratto, 2003). The last objection has been frequently overlooked by the literature, despite the evidence suggesting that individual characteristics play an important role in the choice of sector in general (Dohmen & Falk, 2011; Dohmen et al., 2011) and, in particular, that prosocial preferences are significantly related to the choice of the mission sector (Serra et al., 2011). Given the established literature on the crowding out effect (e.g. see Gneezy & Rustichini, 2000b, 2000a) and mission-oriented effort provision (e.g. see Ashraf, Bandiera, Davenport, & Lee, 2020; Banuri & Keefer, 2016), it is reasonable to think that the use of performance-based incentives might have a negative impact on the amount of effort exerted and the average prosocial motivation of the applicant's pool in the prosocial sector. The effect is particularly important for all the mission-oriented job positions where monitoring is not possible, or it is too costly, and service provision has to partly rely on the worker's intrinsic motivation to reach its recipients. One such instance is described in Deserranno (2019), which looks at the sorting effect of financial incentives in the context of a health-promoter position in Uganda. The author finds that higher expected earnings increase the likelihood of filling the vacancy but discourage more prosocial individuals from applying.

This paper aims to examine how self-selection into different sectors is influenced by high-powered incentives and how they impact performance in a laboratory setting. One of the benefits of using a laboratory setting is that it provides information that would not typically

[1] The definition of mission-driven (or mission-oriented) organisations is based on Besley and Ghatak (2005). The authors divide the economy between the mission-oriented and profit-oriented sectors. In general, the organisations within the mission-oriented sector focus their scope around a mission, i.e., producing a collective good (e.g. non-governmental organisations [NGOs]).

[2] For instance, a series of countries and cities have recently introduced performance-based payments for teachers. The list includes Israel (2000), Denver (2006) and Houston (2006).

be observed in a naturally occurring environment, such as risk and prosocial preferences. I address the following questions: What role do individual characteristics play, and how does the power of the incentives affect sorting into different sectors? Finally, do high-powered incentives lead to higher productivity in the mission-oriented context?

To answer these questions, I employ an experiment divided between an online and lab-based stage. In the online stage, I elicit a vast array of preferences and characteristics with the use of a series of validated survey measures meant to elicit subject's risk preferences, prosocial preferences (trust, altruism, positive reciprocity, and negative reciprocity), patience, self-control, and social desirability (Falk, Becker, Dohmen, Huffman, & Sunde, 2016; Stöber, 2001; Tangney, Baumeister, & Boone, 2004). I then invite respondents to a seemingly unrelated lab experiment, where they are asked to exert effort to gain money. The real-effort task consists of a matrix-solution exercise, where the objective is to add the highest numbers contained in two matrices.

Subjects can choose to work in one of two possible 'sectors': a profit-oriented and a mission-oriented - charity - sector. If subjects decide to work for the profit sector, they generate money for themselves with each additional matrix solved. When choosing the mission-oriented sector, with each additional matrix solved, the subjects generate money that is donated to a charity. The payment scheme associated with the profit sector is always high-powered (i.e. performance-based); a bonus payment is added to the subject's final payoff once they reach a given threshold. In contrast, the incentive scheme provided in the mission-oriented sector varies depending on the treatment; it can either be low-powered (i.e. independent from performance), high-powered, or chosen by the participants (they face a choice between high-powered and low-powered incentives). These are meant to capture different set-ups encountered in the labour market and isolate the effect of the incentive type on the sector's choice.

I consider the treatment where subjects choose between a high-powered profit sector and a low-powered mission sector to be the 'standard setting', as it is the one most commonly encountered in reality. The case where some firms offer low-powered incentives, and others switch to a high-powered one is meant to be captured by the treatment where subjects can choose the power of the incentive within the mission sector. Finally, the treatment where subjects face a high-powered incentive in both sectors represents an extreme case meant to examine what would happen if all firms within the mission sector switched to a high-powered

incentive.

In the standard setting, I find that subjects sort into the mission sector positively based on altruism and negatively based on productivity. I also find a significant gap between sectors in terms of productivity, where subjects solved significantly fewer matrices in the mission one. When participants are given the option to choose the incentive type within the mission sector, those sorting into the low-powered incentive are similar to those selecting it under the standard setting – that is, more altruistic but less productive than those choosing the profit sector. However, I also find that those sorting into the mission sector under the high-powered incentives are just as productive as those that choose the profit sector but are more altruistic. It is reasonable to assume that employers in the mission sector, would want to maximise the number of workers that are both productive and altruistic. The option to choose the incentive type seems to be the most promising venue to reach this goal.

When participants are faced with the same, performance-based incentive in both sectors, my results suggest that altruism is no longer a significant predictor of choice between sectors, but the productivity gap between them disappears. In other words, once the entire sector switches to the high-powered incentive, the productivity gain must be weighed against the potential loss in terms of the average altruism of the applicant's pool. Finally, the use of high-powered incentives results in a reduction in the number of subjects applying to the mission sector. In turn, this leads to a lower level of charitable donations, despite the increase in productivity due to the incentive effect.

The evidence I present in this paper is of particular interest to policymakers, especially given the recent rise in the number of programmes aiming at increasing the productivity in the public/mission sector in developing countries with the use of performance-based incentives (e.g. Binyaruka et al., 2015; Muralidharan & Sundararaman, 2011; Duflo et al., 2012). My results replicate the problem with the use of low-powered incentives in the prosocial sector, as they attract less productive individuals. They also show that if the option is a mere choice between sectors with the same high-powered incentive, it would generate a trade-off between lower altruism and higher productivity, possibly arising because they make the extrinsic reward more salient. Another potential explanation relies on the idea that performance-based incentives can be viewed as an indication of a lack of trust in the worker; the subjects could perceive that they are not trusted enough to rely solely on their intrinsic motivation. Even though it is not the charity choosing the incentive in this context, it is predetermined by the

experimenter. Overall, the result is that fewer subjects enter the mission sector - and those who enter are not more prosocial than those who do not enter. This might not sound like a big problem at first, but it is important to consider the type of service provided by the mission sector and how much they rely on prosocial motivation for their delivery. The most significant advantage from a mission-driven firm's perspective would be to switch to a high-powered incentive, as long as other firms do not do the same thing. The gain realised by the individual firm depends on how the rest of the firms will behave; as more switch to a high-powered incentive, the gain gets lower. Nonetheless, it is very reasonable to expect that not all of them will be able to change the incentive scheme, given that for some of them, tying the payment to a measurable output would not be possible. In my setting, the participants cannot choose between several companies as would typically happen in the labour market. Nonetheless, by using one representative firm, I can still reach conclusions that can generalise to a wider framework.

The paper proceeds as follows. **Section 2** discusses the related literature and my contribution to it. **Section 3** outlines the experimental design and **Section 4** provides the data analysis. **Section 5** concludes the paper.

2 Related Literature

This study on the effect of high-powered incentives in the mission sector complements and contributes to three main strands of the literature, namely sorting, motivation crowding out, and the impact of incentives on mission-oriented effort provision. It complements the literature on sorting by examining how incentives interact with the choice of sectors. The existing literature suggests that incentive schemes affect not only productivity but also how individuals self-select into different sectors and occupations, which attract different types of subjects based on their characteristics. For example, when given a choice between variable (performance-based) and fixed payment, more productive workers systematically self-select into variable payment schemes (Cadsby, Song, & Tapon, 2007; Dohmen & Falk, 2011; Eriksson & Villeval, 2008). Other individual characteristics, such as risk and prosocial preferences, are significantly related to workplace selection. Prosocial preferences, in particular, have been found to be an important driver of workers' self-selection into the non-profit sector both in the field (Serra et al., 2011; Lagarde & Blaauw, 2014) and in lab

experiments (Banuri & Keefer, 2013). Mission-driven organisations benefit from hiring intrinsically motivated workers, as they tend to exert more effort and require fewer extrinsic incentives than self-interested ones. The two topics have been studied separately before, but this paper is among the first ones to study their interaction, applying the findings on the effect of performance payments to sorting effects in the mission-oriented context.

Previous literature on mission-oriented effort provision has mostly focused on examining experimentally (Banuri & Keefer, 2013; Bellé & Cantarelli, 2015; Dal Bó, Finan, & Rossi, 2013; Deserranno, 2019) and theoretically (Barigozzi, Burani, & Raggi, 2018; Delfgaauw & Dur, 2007; Ashraf et al., 2020) how the level of the incentive affects selection into the public/mission-driven sector, i.e. how increasing or decreasing fixed payments affects the productivity and prosociality in the applicant's pool and, in some cases, service provision. When it comes to the effect on prosociality and public service motivation, the experimental literature is divided. Some papers, such as Dal Bó et al. (2013), found that increasing prospective salaries increases the quality of the applicants (as measured by their IQ) without decreasing their public service motivation (as measured by a questionnaire). On a similar note, Ashraf et al. (2020) experimentally varies the extent to which career benefits are stressed in a recruitment drive for nurses in Zambia, finding that the trade-off between talented and prosocial workers is only present at low levels of talent. Conversely, Deserranno (2019) showed how offering higher financial incentives for a job with a strong prosocial component, such as health worker, changes agents' perceptions of the job. The reason being that more lucrative positions are perceived as involving lower positive externalities for the community and therefore fail to recruit the most prosocially motivated individuals. Also, Banuri and Keefer (2016) analysed the relationship between prosocial motivation, as measured by a modified dictator game, and effort in a lab setting using Indonesian students destined for the profit and public sectors. Their results suggest that increasing salaries in the prosocial sector leads less prosocially motivated individuals to apply. One explanation for this discrepancy within the literature is that each paper focuses on a different aspect and uses a different measure of prosociality. It is within this literature that lies the main contribution of this paper. In particular, this is one of the first studies examining how the use of high-powered incentives, as opposed to merely varying the level of fixed incentives, affects effort and selection within the mission sector.

Given its focus on the effect of incentive types, this paper also relates to the literature on

the crowding out effect, which suggests that incentives shape outcomes, but not necessarily in the desired direction (for a summary, see Gneezy, Meier, & Rey-Biel, 2011). When incentive schemes are not carefully designed to match the underlying motivation of individuals, they can, for example, change the perception of the incentivised action. Bénabou and Tirole (2006) suggested that extrinsic incentives (e.g. money) can crowd out intrinsic motivation by introducing doubts about the reasons behind a given prosocial action. Famously, research has shown that introducing monetary incentives for blood donation reduces both the quantity (Mellström & Johannesson, 2008) and the quality (Eastlund, 1998) of the donated blood. The main takeaway being that extrinsic incentives have the potential to undermine the signalling value of a given action.

Another way in which extrinsic incentives can affect prosocial motivation is through trust. Stronger monetary incentives can signal mistrust. Monitoring systems and pecuniary incentives have the potential to crowd out trust and negatively affect performance (see, e.g. Falk & Kosfeld, 2006), that is especially when monitoring is above a certain threshold (Dickinson & Villeval, 2008) or it is perceived as revealing pessimism about the social norm (Sliwka, 2007). When comparing the behaviour of students and that of CEOs, Fehr and List (2004) found that CEOs are more trusting than students and that incentives based on explicit threats to penalise shirking backfire, because they induce less trustworthy behaviour. In other words, their findings suggest that trust induces trustworthiness. In summary, there are two main mechanisms by which the use of high-powered incentives has the potential to impact performance and prosocial motivation. The first is by shifting the focus and meaning of effort from a prosocial goal to a market/productivity-oriented goal. The second is by introducing a sense of distrust in the subjects, who may believe that the employer does not trust their motivation enough to rely on it to provide a given good or service.

To the best of my knowledge, the paper that most closely resembles this study is Jones, Tonin, and Vlassopoulos (2018), which examines the effect of high-powered incentives in the public/mission sector. However, that paper focuses mainly on the impact of high-powered incentives on productivity, whereas my study is primarily concerned with the effect on selection. Additionally, a series of design features separate their paper from mine. The most important ones are the incentive structure and the possibility to choose between high-powered and low-powered incentives within one of the treatments. Another closely related paper, Green and Kloosterman (2019), examines theoretically and experimentally the effect

of different types of incentives in credence goods markets, and finds a trade-off between labour supply and service quality for the mission firms. In particular, their results suggest that as performance-based incentives attract more profit-oriented individuals, they decrease the quality of the services provided. Notably, their paper does not consider individual characteristics and how they affect sorting between different incentive types.

3 Experimental Design

The experiment is divided into two parts, an online stage and a lab-based experiment. Subjects were not explicitly told about this structure and took the experiment as two separate studies.[3] The online stage consists of a questionnaire, which is sent to the prospective participants at least two weeks before the lab experiment. The questionnaire aims at gathering data on socioeconomic characteristics (age, gender, nationality, etc.) and provide a survey measure of the respondents' risk aversion, discounting, trust, altruism, positive reciprocity, negative reciprocity, social desirability, and self-control. The payment of the first part was based on a lottery in which five participants - among all those that completed the study - were randomly chosen and paid £50 independently from their answers.

Only those that replied to the online stage were then invited to the lab experiment. The structure of this second part of the study is summarised in Figure 1; it consists of three stages. Subjects were informed that they would go through the different stages but were not told what each entailed before completing the previous ones. Participants were presented with two 6X6 matrices in which each cell is filled with positive integers from 1 to 99, positioned next to each other.[4] Solving a matrix involves summing the highest number contained in the left matrix with the highest number contained in the right matrix, reporting it in the input box, and clicking the Submit button. After the correct answer is provided, the next pair of matrices appears until the time allocated to the task is elapsed. The participants were informed about the remaining time and the number of matrices correctly solved by two counters positioned below the matrices. There was no penalty for incorrect answers. However, subjects were not allowed to proceed to the next set of matrices until they answered correctly.

[3] Detailed experimental instructions can be found in the Instructions section at the end of the thesis.
[4] A similar task was previously used in Fallucchi, Nosenzo, and Reuben (2020).

FIGURE 1. Summary of the structure of the lab-based part of the experiment

Stage 1	Stage 2	Stage 3
Productivity Measure	**Sorting**	**Working Time**
5 minutes	High-Low Profit-High power vs Mission-Low power	**10 minutes**
Matrix solution task 20p per correct answer + Effort-related questions	High-Choice Profit-High power vs Mission-Low power vs Mission-High power	Matrix solution task + Effort-related questions
	High-High Profit-High power vs Mission-High power	

Independently from the treatment, all participants went through the first stage to familiarise themselves with the task and, more importantly, determine their baseline productivity. Subjects were asked to solve the real-effort task and were paid a £0.20 piece rate for each correct answer. In the second stage, subjects were asked to choose one of two possible 'work sectors': either a mission sector (where the output of their work generated profit for a charity) or a profit sector (where the work generated profit for themselves). The type of incentive faced in each sector changed depending on the treatment. Once subjects made their choice, they moved to the third stage, where they worked for 10 minutes in the sector of their choice under the corresponding incentive scheme. A final demographics questionnaire ended the study.

3.1 Online Stage

In the online stage, I adopted the survey module of Falk et al. (2016), which allowed me to elicit risk, time, and prosocial preferences (trust, altruism, positive reciprocity, negative reciprocity). The authors selected the items as a result of a horse race between questions to choose the ones that were jointly the best predictors of behaviour in a series of laboratory

experiments and are, therefore, experimentally validated. The reasons behind this choice were mostly practical. One benefit of adopting the Falk et al. (2016) survey items is that they allow measuring a broader spectrum of preferences, which would have only been possible using many different tasks. In the behavioural economics field, incentivised measures are considered a golden standard to elicit subjects' preferences, particularly risk and prosocial preferences. Although, in recent years there has been a rise in the number of experimentally validated survey instruments developed (e.g. see Dohmen et al., 2011; Falk et al., 2016; Fallucchi et al., 2020; Buser, Niederle, & Oosterbeek, 2021) and papers applying these instruments to large-scale studies (Falk et al., 2018; Falk & Hermle, 2018; Becker, Enke, & Falk, 2020).

Self-control was evaluated using the Brief Self-Control Scale (Tangney et al., 2004), one of the most widely adopted instruments in the psychology field to measure trait self-control. To estimate the subjects' need for social approval, I employed the Social Desirability Scale (SDS-17; Stöber, 2001). The scale is constituted by the sum of answers to a series of 16 items measuring the extent to which subjects over-report socially desirable behaviours and under-report socially undesirable ones. Examined for both convergent and discriminant validity, the SDS-17 scores were found to significantly correlate with other social desirability measures and predict behaviour in a series of settings (Stöber, 2001).

3.2 Lab Experiment

Subjects were randomly assigned to one of three treatments: *High-Low* (high-powered profit sector vs. low-powered mission sector), *High-Choice* (high-powered profit sector vs. chosen power of the incentive in the mission sector), and *High-High* (high-powered profit sector vs. choice between high-powered and low-powered incentive in the mission sector). The High-Low treatment was meant to mimic the standard setting where the private/profit-oriented sector pays a performance-based salary, made of a piece-rate and a threshold component, and the public/mission-oriented sector pays a fixed salary. The structure of the low-powered incentive aimed to capture the idea that those working in the mission-oriented sector are often paid based on a fixed salary, but the additional effort they exert benefits the charity they work for and society in general. It consists of a fixed component and a piece rate. The participants earn £3 regardless of the number of matrices they solve. However, each matrix correctly solved generates £0.05, multiplied by two and donated to their chosen charity. Upon

selecting the mission sector, independently from the treatment, participants were given a list of eight charities to which they could donate the money they earn in the experiment. Each organisation in the list focuses on a different mission (e.g. Amnesty International for human rights, Red Cross for health, etc.) to allow the matching between the subject's preferred mission and the organisation's objectives. In the profit sector, the high-powered incentive combines a piece-rate with a threshold bonus payment. Participants who solve at least 17 matrices were paid a £5 bonus on top of the final earnings; otherwise, they were paid based on a piece-rate of £0.05 for each matrix solved.

The High-Choice treatment allows subjects to choose the type of incentive they want to face in the mission sector. It represents a transition case, where some mission-oriented firms adopt a high-powered incentive, whereas others retain a low-powered one. This is possibly one of the most realistic treatments. It partly captures the labour market's heterogeneity, where subjects can choose between firms offering the same service but paying different types of incentives. The main objective was to examine if and how workers self-select into different types of payment and whether it leads to a difference in productivity within the mission-oriented sector. We can also think of this setting as a case where there are two mission-driven organisations, one that offers a low-powered incentive and the other one that offers a high-powered incentive.

The High-High treatment examines the case in which both sectors pay high-powered incentives. It represents an extreme circumstance, which would be challenging to examine outside the lab, as it captures the consequences of the entire mission sector shifting to performance-based incentives. In other words, for this treatment, I decided to eliminate the heterogeneity characterising High-Choice and describe a situation where the labour market is only populated by firms offering a high-powered incentive. The incentive scheme is the same, but depending on whether the subject chose to work in the profit or the mission sector, they either got the piece-rate themselves or gave it to the selected charity multiplied by two. That is, those subjects that chose to work in the mission sector but did not reach the threshold gained nothing for themselves. Whereas those who did not reach the 17 matrices threshold in the profit sector still got to keep the piece-rate.

3.3 Procedures

The experiment was programmed on the web-based platform LIONESS (Giamattei, Yahosseini, Gächter, & Molleman, 2020). The data were gathered in two waves: one in May 2019 and the next in November 2019. The structure and running procedures were kept constant across waves.

The invitations to the online stage were sent to the prospective participants at least two weeks before the lab-based stage. All the subjects inscribed to the subjects pool of the University of Nottingham were contacted through the ORSEE (Greiner, 2015) recruitment platform and asked to participate in an online experiment. At the end of the two days given to start and complete the survey, I gathered 1,298 answers between the two waves. I extracted and successfully contacted all the lottery winners and paid them £50. Those who replied to the first stage were then invited to the second stage; that is, the invitation to the lab experiment was conditional on having completed the first online part of the experiment.

For the lab-based part of the experiment, the general instructions were read aloud at the start of each session. The subjects were made aware from the beginning that the main task would involve adding numbers and were explicitly told that it was not possible to use any helping device (e.g. mobile phones and calculators). During the sessions, one of the experimenters ensured that the rule was followed by walking around and checking the cubicles at regular time intervals. The specific instructions of the experiment were presented on the computer screen. I ran 46 sessions, with an average of nine participants. In total, 372 subjects participated in the lab experiment, which took between 20 and 30 minutes to complete. The subjects were all students at the University of Nottingham. The average earnings were about £6.50.

4 Results

4.1 Behavioural Measures

About 28% of those who replied to the first stage signed up for the second stage. Given the high attrition rate between the different stages of the experiment, the first step is to examine whether there is any evidence of selection based on the characteristics elicited in the online part. Figure A.1 reports the standardised marginal coefficients plot for a probit regression

where the dependent variable equals one if the subject inscribed to the second stage and zero otherwise. The results show some evidence of selection in terms of altruism, income, and age; that is, older, more altruistic and lower-income individuals were more likely to inscribe to the lab stage. Other weakly significant independent variables are patience, positive and negative reciprocity. However, given that the magnitudes are mostly very small, there is limited evidence of selection on the attributes of interest.

A statistical summary of the participant's characteristics is contained in Table A.1, along with the results of the balancing tests. The majority of the participants in the lab stage were females, with an average age of 22 years. Given that the real-effort task relies on basic mathematical skills, I also asked participants about their confidence level in maths and find that the average level of self-reported mathematical skills is evenly distributed across treatments. When it comes to the main variables of interest, risk-taking is the only variable that is not evenly distributed (Kruskal Wallis test, $\chi^2(2) = 6.026$, $p = 0.0463$, $N = 371$). In particular, in the High-Low treatment, the average risk measure is slightly higher than in the other two treatment conditions (9.44 vs. 8.43 vs. 8.64). Additionally, I report in Figure A.2 a series of histograms showing the standardised distribution of the main variables. The only two variables that are normally distributed are negative reciprocity and trust (Shapiro-Wilk test; negative reciprocity: $z = 0.998$, $p = 0.159$, $N = 371$; trust: $z = -2.953$, $p = 0.998$, $N = 371$). Most of the regressions contained in the following use standardised independent variables. One thing to keep in mind is that standardisation does not change the underlying distribution of the variables. In this case, not all variables are normally distributed, so even when standardising, it is not possible to make direct comparisons in terms of magnitudes between coefficients.

Table A.2 reports the results of the pairwise correlation analysis. The variable which is more often significantly correlated with the others is social desirability. In particular, the highest correlation is found to be the one between social desirability and self-control (0.41; Spearman rank correlation, $p < 0.01$). Altruism is significantly related to many independent variables, but the correlations are mostly small in magnitude. It is important to note that altruism and productivity are not significantly correlated. In other words, at least at the initial stage of analysis, there is no evidence of a relationship between altruism and the number of matrices solved in the working stage.

4.2 Selection into the mission-oriented sector

In the following, I analyse the determinants involved in the choice of the mission sector and how they vary depending on the treatment. Previous research suggests that more productive workers systematically self-select into the performance payment scheme (Dohmen & Falk, 2011; Lazear, 2000). The complication, in this case, is due to the interaction between productivity and prosocial motivation. When faced with the choice between high-profit and low-mission, the choice depends not only on the first-stage productivity but also on prosocial preferences, altruism in particular. When subjects are not very altruistic, the data should point in the same direction as previous research. In essence, low-productivity types should choose the mission sector, and high-productivity types should choose the profit sector. On the other hand, altruistic subjects should be more inclined to select the mission sector in general. However, what altruistic types with high productivity will do is not easy to determine a priori, especially in the High-Low treatment, where they face a trade-off between altruism and productivity. In the High-Choice treatment, they should choose the mission sector under the high-powered incentive, thus separating from altruists with low productivity. Whereas, in the High-High treatment, these subjects should, in principle, choose the mission sector.

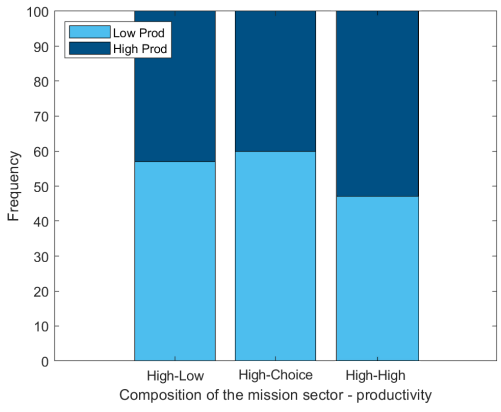
In the High-Low treatment, 36% of the participants choose the profit sector, and 64% decided to work in the mission sector. In the High-Choice treatment, 71% of the subjects choose to work for the mission sector (50% under the low-powered incentive and 21% under the high-powered incentive) and the remaining 29% chose the profit sector. Finally, for the High-High treatment, the shares inverted, as I observe 62% of participants selecting the profit over the mission sector.

Taking a step forward in the analysis, Figure 2 shows the raw change in the composition of the mission sector between treatments. The percentages are calculated over the number of subjects that chose the mission sector. The sample is divided between low- and high-altruism types and low- and high-productivity types depending on whether the subjects were above or below, respectively, the median of the altruism measure and the median of the productivity measure, which consists of the number of matrices solved in the initial 5-minute period. Those categorised as low-productivity solved 7 matrices or fewer in the first stage, thus having a low probability of reaching the threshold of 17 matrices in the working period.[5]

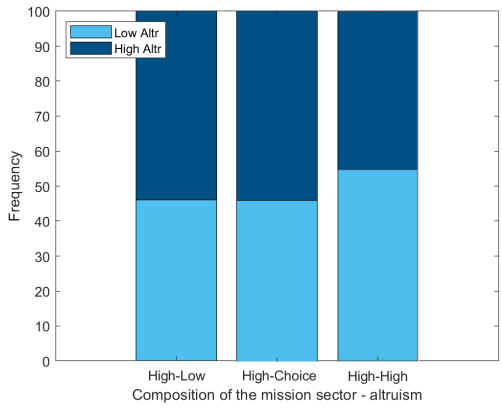
[5]The first stage lasted 5 minutes, whereas the working stage lasted 10 minutes. Multiplying the performance in the first stage by two, I obtain an approximate estimate of whether subjects are likely to reach the 17 matrices

Those classified as high productivity solved more than 7 matrices in the first stage. Starting with productivity, we can see that the fraction of high productivity subjects in the mission sector is similar in High-Low and High Choice, while it is higher in High-High.

FIGURE 2. Change in the composition of the mission sector between treatments



(A) Productivity



(B) Altruism

When it comes to altruism, there is a similar pattern as the one noted earlier. We can observe a lower share of subjects categorised as high altruism choosing the mission sector in High-High compared to the rest of the treatments. Even though for High-High, the choice between the mission and the profit sector should be based only on altruism, as both sectors pay the same type of incentive. If altruism is a fixed trait, independent of the incentive schemes offered, it should not depend on the type of incentive offered in the mission sector. That is, I should observe that it predicts the choice of the mission sector even under the high-powered incentive; however, this is not what I find.

Taken together, these preliminary findings show that the type of incentive faced in the mission sector does impact its composition both in terms of altruism and productivity. Additional details are shown in Table A.3, which reports how many of those subjects who chose the mission sector fall into each category.

In the following, I examine how sorting is affected by the option to choose the type of incentive in the mission sector. Those who were assigned to the High-Choice treatment were able to decide between working in the profit sector under a high-powered incentive or selecting the power of the incentive they wanted to work for in the mission sector; once they chose the mission sector, they could decide between a fixed payment and a performance-based

threshold in the working stage.

payment. We can think about this setting as mimicking a situation where subjects have the option to choose between two firms within the mission sector, one offering a low-powered incentive and the other offering a high-powered incentive. Alternatively, we could also think about it as the same firm offering the option to choose the payment type.

Table 1 shows the relative risk ratios (relative probabilities) for a multinominal logit model, the same analysis but for the marginal effects is contained in Table A.5. In particular, Column (1) and (2) report the relative probabilities comparing $Profit$ and $Mission_H$ with $Mission_L$. Column (3) reports the relative probabilities comparing $Profit$ with $Mission_H$. A risk ratio greater than one indicates the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the reference group increases as the variable increases. That is, the comparison outcome is more likely. A risk ratio that is smaller than one indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group decreases as the variable increases, such that the outcome is more likely to be in the reference group.

Starting from Column (1), the most significant difference in terms of risk ratios is the one related to the productivity measure. An increase of one standard deviation on the productivity scale would result in the rise of the relative risk of choosing $Profit$ over $Mission_L$ by a factor of 3.337, all other things equal. This finding is in line with the idea that more productive subjects are more likely to work under the high-powered incentive. When focusing on altruism, we can see that it is a significant determinant of behaviour. In particular, an increase of one standard deviation on the altruism scale leads to a decrease in the relative risk of choosing the profit sector over the mission sector under the low powered incentive by a factor of 0.529. More generally, we can say that an increase in the altruism measure would lead a participant to be more likely to prefer $Mission_L$ over $Profit$. A higher score on the social desirability scale, on the other hand, would dictate the choice of the profit sector over the mission sector.

Moving to Column (2), we can see that productivity is, again, a strongly significant determinant of behaviour within the same sector, as higher productivity increases the relative likelihood of choosing the high-powered incentive over the low-powered incentive. It is worth noting that the effect is even more prominent than in the previous column. The relative risk for altruism is oriented towards the comparison group. Still, it is both small in magnitude and not significant, meaning that, within the same sector, I do not find any evidence of sorting

	(1) $Profit$ vs $Mission_L$	(2) $Mission_H$ vs $Mission_L$	(3) $Profit$ vs $Mission_H$
Productivity	3.337***	4.123***	0.809
	(1.092)	(1.369)	(0.219)
Altruism	0.529*	1.171	0.451*
	(0.195)	(0.346)	(0.189)
Risk	1.369	1.523	0.898
	(0.352)	(0.507)	(0.316)
Patience	0.802	1.276	0.629
	(0.234)	(0.343)	(0.201)
P. reciprocity	1.105	1.162	0.952
	(0.265)	(0.347)	(0.314)
N. reciprocity	1.155	0.904	1.277
	(0.302)	(0.278)	(0.395)
Trust	1.059	0.746	1.420
	(0.280)	(0.194)	(0.438)
Self-control	0.896	1.638	0.547*
	(0.215)	(0.530)	(0.188)
Social desirability	1.939**	0.941	2.062**
	(0.526)	(0.311)	(0.756)
Controls	Yes	Yes	Yes
Pseudo R^2	0.064	0.064	0.167
Observations	124	124	124

Notes: Multinomial logit estimates. Relative risk ratios reported; robust standard errors between parenthesis. The baseline for the first two columns is $Mission_L$ and for the last Column is $Mission_H$. The controls include: Age, Female and Income.

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

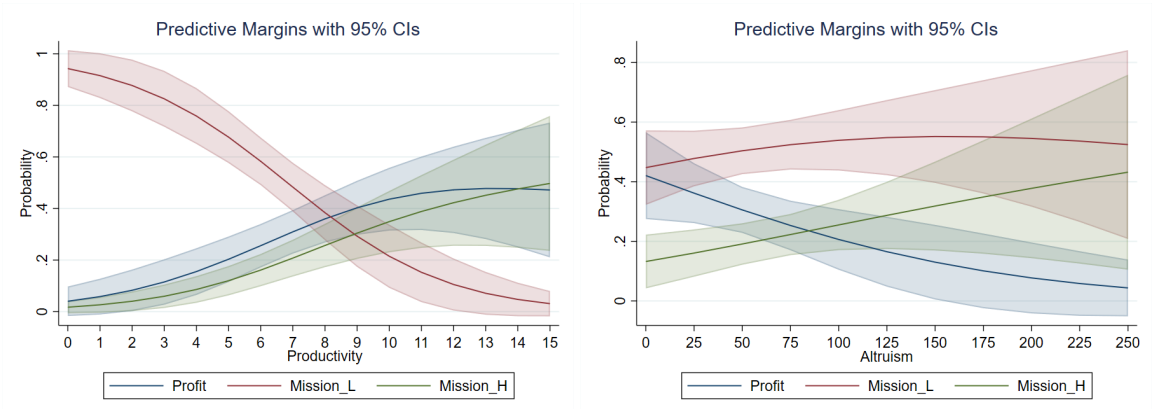
TABLE 1. Determinants of the choice of sector and incentive type for the High-Choice treatment

based on altruism.

Finally, Column (3) reports the comparison between $Profit$ and $Mission_H$. The productivity measure is not significantly related to the choice of either the profit or the mission sector under the high-powered incentive. It is, however, oriented towards the latter. Altruism is weakly related to the choice of the mission sector under the high powered incentives over the profit sector. The other two significant determinants are self-control and social desirability, which go in opposite directions. Whereas higher self-control increases the relative likelihood of choosing the mission over the profit sector, social desirability increases the probability of selecting the profit sector. The latter result is not easy to explain and yet coherent with

the results found in Column (1) and the previous regressions. Given the previous evidence suggesting a negative relationship between social desirability and contributions in a public goods game (e.g., Fleming & Zizzo, 2011), this last result sounds counter-intuitive but not particularly surprising.

FIGURE 3. Likelihood of choosing each option based on productivity and altruism



(A) Productivity

(B) Altruism

Figure 3a shows the predictive margins of productivity for all the outcome levels. The lines representing the likelihood of the choice of $Profit$ and $Mission_H$ have a similar trend. However, the likelihood of choosing $Mission_L$ goes in the opposite direction. For high productivity levels, the probability of selecting $Profit$ or $Mission_H$ is higher than the one of choosing $Mission_L$. The likelihood of selecting the mission sector under the high-powered incentives increases linearly with altruism, as seen in Figure 3b. On the other hand, when incentives are low-powered, the relationship looks non-linear, and the probability of choosing that option over the others is higher for all levels of altruism, including low ones. It is worth noting that the probabilities of choosing $Mission_L$ and $Profit$ start very close and progressively distance themselves as altruism increases. The effect is due mainly to the likelihood of selecting $Profit$ decreasing with altruism.

Overall, the results suggest that productivity is an important determinant of choice both within and between sectors. In contrast, altruism is an important determinant of choice between but not within sectors, even though the first is admittedly stronger and more significant than the second. More productive subjects are more likely to choose the high-powered incentives, and more altruistic ones are more likely to select the mission sector.

Tables A.6, A.7 and A.8, report the marginal effects for a series of logit regressions where

	(1) High-Low	(2) High-Choice	(3) High-High
Productivity	-0.101**	-0.123***	0.013
	(0.048)	(0.040)	(0.046)
Altruism	0.145**	0.132**	0.039
	(0.057)	(0.063)	(0.046)
Risk	-0.041	-0.030	-0.018
	(0.046)	(0.043)	(0.051)
Patience	0.041	0.061	-0.040
	(0.048)	(0.047)	(0.044)
P. reciprocity	-0.001	-0.001	0.068
	(0.055)	(0.041)	(0.051)
N. reciprocity	-0.136**	-0.026	-0.092*
	(0.064)	(0.039)	(0.049)
Trust	-0.067	-0.033	-0.090**
	(0.051)	(0.044)	(0.046)
Self control	-0.054	0.050	-0.018
	(0.058)	(0.040)	(0.061)
Social desirability	-0.111**	-0.121***	0.072
	(0.054)	(0.045)	(0.052)
Controls	Yes	Yes	Yes
Pseudo R-squared	0.206	0.081	0.144
Observations	122	124	125

Notes: Logit estimates. Marginal effects (evaluated at the mean of independent variables) reported; robust standard errors between parenthesis. The outcome variable takes value one if the subject chose the mission sector. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

TABLE 2. Determinants of the choice of the mission sector

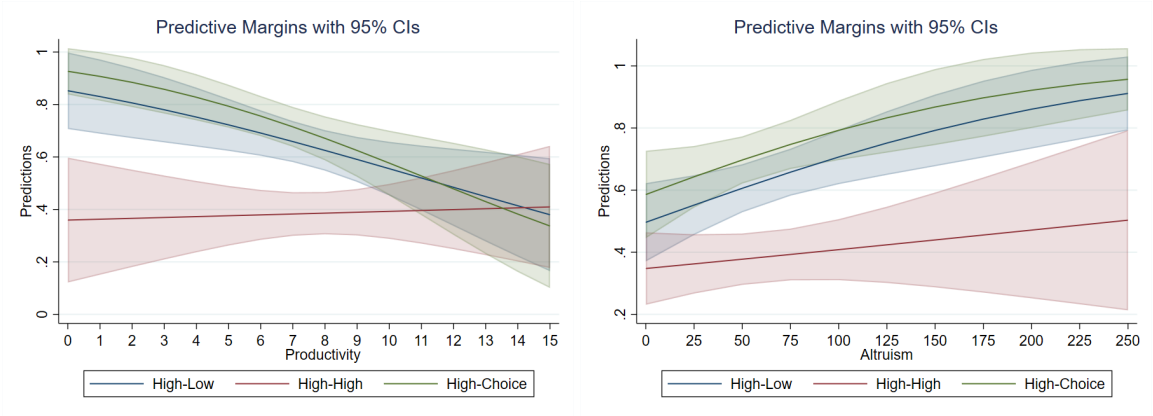
the dependent variable is equal to one if the subject chose the mission sector, respectively, in the High-Low, High-Choice and High-High treatment. The last columns of each regression are reported in Table 2. When it comes to the High-Choice treatment, I collapsed the two incentives options to examine the overall determinants of the choice of the mission sector.

In the High-Low treatment, productivity is negatively and significantly related to the choice of the mission sector, as subjects with low productivity can choose the mission sector to avoid the high-powered incentive. In particular, one standard deviation increase in the number of matrices solved in the first stage reduces the likelihood of choosing the mission sector by

approximately 10%.

In the High-Choice treatment, the relationship between the choice of the mission sector and the productivity measure is again negative and significant. Finally, for the High-High treatment, lower-productivity subjects cannot opt-out of the performance payment by selecting the mission sector as they both offer the same, high-powered, incentive. In fact, in that case, productivity is not a significant predictor of the choice of the mission sector. The findings relating to productivity and the choice of the mission sector across treatments are summarised in Figure 4a.

FIGURE 4. Likelihood of choosing the mission sector based on productivity and altruism



(A) Productivity (B) Altruism

A significant relationship exists between altruism and the choice of the mission sector for the High-Low and High-Choice but not for the High-High treatment. As shown in Table A.8, I find that independently from the specification, altruism does not significantly predict the choice of the mission sector for the High-High treatment. This is particularly surprising. In principle, the only reason subjects should be drawn to the mission sector in this context is altruism, given that the sectors pay the same type of incentive.

Figure 4b summarises my findings regarding altruism; there, we can see a strong and positive relationship between altruism and the choice of the mission sector in the High-Low and High-Choice treatment. As altruism increases, so does the probability of choosing the mission sector. For the High-High treatment, the relationship is positive, but the line is less steep. The gap between the lines is initially narrow but then widens as the level of altruism increases. In other words, for the same level of altruism, the likelihood of choosing the mission sector when incentives are low-powered or chosen by the participants is higher.

Previous literature found risk attitudes to be a significant determinant of the choice of performance payments (see, e.g. Dohmen et al., 2011). However, in this instance, I do not find a link between the choice of sector and risk preferences. In other words, of the 10 variables included in the regression, only two are significant determinants of the mission sector choice in the High-High treatment. These variables are negative reciprocity and trust, which negatively affect the likelihood of choosing the mission sector. I find negative reciprocity to be the only predictor of the choice of sector that is significant across the High-Low and High-High treatment. The sign suggests that those higher on the negative reciprocity scale are less likely to choose the mission sector. Interestingly, trust is observed as negatively related to the choice of the mission sector in the High-High treatment. As mentioned earlier, there are different channels by which prosocial motivation can be affected by extrinsic incentives. These results seem to suggest that part of the effect might be mediated by a reduction in the level of trust. The reason is that the use of high-powered incentives in the mission sector could signal a lack of trust towards the subjects, instilling the feeling that they are not trusted to put enough effort without it. Performance-based incentives might, therefore, signal distrust, resulting in the negative effect we observe. However, I do not find the average trust level to be different between treatments (Kruskal-Wallis test, $\chi^2(2) = 0.726$, $p = 0.695$, $N = 371$). Finally, similarly to the previous table, I find that social desirability is negatively and significantly related to the choice of the mission sector in High-Low and High-Choice.

To summarise, when subjects are faced with the standard setting - high-powered profit sector versus low-powered mission sector - my findings are in line with the previous literature suggesting that more altruistic, but less productive, individuals are more likely to self-select into the prosocial sector (e.g. Serra et al., 2011). When focusing on the High-Choice treatment, we can see that when subjects are given the option to choose the power of the incentive in the mission sector, the low-powered payment type prevails. These findings corroborate those reported in the High-Low treatment as more prosocial but less productive types choose the mission sector, mostly selecting into the low-powered incentive. Finally, when moving to the case in which both sectors pay a high-powered incentive, I find that altruism is no longer a significant determinant of the choice of the mission sector. In other words, when only offering payments based on performance, the link between altruism and the mission sector choice becomes weaker, suggesting a potential trade-off between altruism

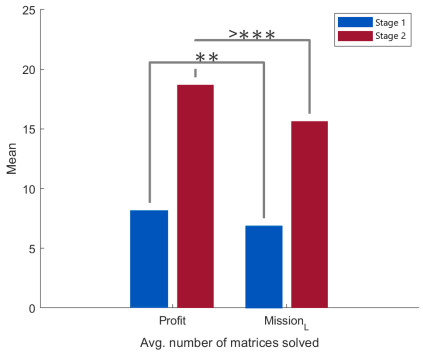
and productivity. One of the possible explanations for my findings regarding altruism relies on the literature on the effect of higher monetary incentives in the mission sector (Banuri & Keefer, 2016; Deserranno, 2019). The idea is that both increasing financial incentives and using high-powered incentives effectively shift the focus from the charitable component to the monetary one. This effect is particularly salient in the High-High treatment, as the choice is restricted between two options. This result is very important in light of the goods produced by the prosocial sector and how much they often rely on intrinsic motivation when it comes to their delivery.

4.3 Output

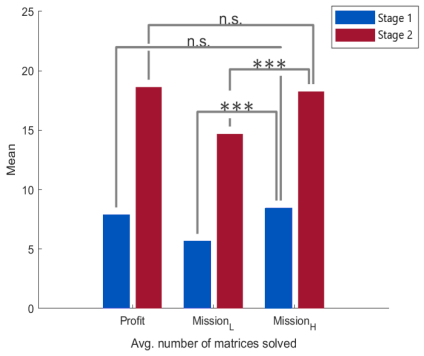
In line with incentive theory, Figure 5a shows that, for the High-Low treatment, productivity was highest in the profit sector (Wilcoxon rank-sum test, z = 3.819, $p = 0.000$, N = 122). Subjects solved on average 18.70 (sd: 4.42) matrix problems, compared to the 15.64 (sd: 3.91) of those who chose the mission sector. There is a high and significant correlation between performance in Stage 1 and Stage 2 – that is, before and after the choice of sector has been made (Spearman rank correlation, $p < 0.01$) – suggesting that the performance measure is significantly related to the performance in the second part of the experiment. In the first stage, all subjects face a piece-rate, so the difference in terms of productivity cannot be explained by a difference in the choice of the incentive.

In the High-Choice treatment, shown in Figure 5b, subjects in the profit sector solved 18.63 (sd: 5.39) matrices, and those in the mission sector solved 14.69 (sd: 4.70) for the low-powered incentive and 18.26 (sd: 4.03) for the high-powered one. Statistical tests confirmed a significant difference between the output in the profit sector and low-powered mission sector (Wilcoxon rank-sum test, $z = 3.716$, $p = 0.002$, $N = 98$) but found no difference between the profit sector and high-powered mission sector (Wilcoxon rank-sum test, $z = -0.100$, $p = 0.924$, $N = 62$).

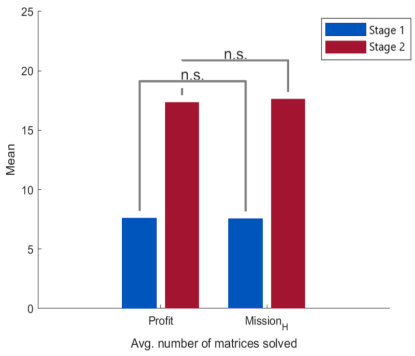
FIGURE 5. Average number of matrices solved in the first and second stage across treatments



(A) High-Low



(B) High-Choice



(C) High-High

Notes: The stars represent the significance levels of the Wilcoxon ranksum tests reported in the text. * p < 0.1, ** p < 0.05, *** p < 0.01

The average obtained by combining both incentives in the mission sector in the second stage is significantly lower when compared to High-High (stage 1: Wilcoxon rank-sum test, $z = 2.220$, $p = 0.026$, $N = 136$; stage 2: Wilcoxon rank-sum test, $z = 1.891$, $p = 0.0586$, $N = 136$) but not when compared to High-Low (stage 1: Wilcoxon rank-sum test, $z = -0.100$, $p = 0.924$, $N = 166$; stage 2: Wilcoxon rank-sum test, $z = 0.099$, $p = 0.921$, $N = 166$).

For the High-High treatment there is no difference in terms of productivity between the two sectors both in the fist (Wilcoxon rank-sum test, $z = 0.164$, $p = 0.871, N = 125$) and in the second stage (Wilcoxon rank-sum test, $z = -0.260$, $p = 0.795$, $N = 125$). In particular, in stage 2, subjects solved on average 17.64 (sd: 5.14) matrices in the mission sector and 17.36 (sd: 5.10) in the profit sector.

	(1) N. solved 2	(2) N. solved 2	(3) N. solved 2
High-Choice	0.062	-0.172	-0.175
	(0.512)	(1.382)	(1.396)
High-High	0.106*	-0.323*	-0.310
	(0.623)	(2.069)	(2.130)
Productivity	0.685***	0.529***	0.509***
	(0.096)	(0.143)	(0.143)
H-C * Productivity		0.260	0.258
		(0.211)	(0.210)
H-H * Productivity		0.505***	0.464**
		(0.263)	(0.270)
Altruism			-0.033
			(0.004)
Risk			0.033
			(0.076)
Patience			-0.014
			(0.028)
P. reciprocity			0.078*
			(0.079)
N. reciprocity			0.040
			(0.142)
Trust			0.028
			(0.098)
Self control			0.049
			(0.027)
Social desirability			-0.019
			(0.120)
Controls	No	No	Yes
R-squared	0.489	0.507	0.539
Observations	214	214	214

Notes: Ordinary least squares (OLS) estimates. Robust standard errors between parenthesis. The outcome variable is the number of matrices solved in the second part of the experiment for the mission sector. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

TABLE 3. Determinants of output in the mission sector

These findings are summarised by Figure A.4, which reports the cumulative distribution of the number of matrices correctly solved in the 10 minutes working time. The Kolmogorov-Smirnov tests finds a strong difference between sectors in the High-Low treatment (D =

0.347, p = 0.002) and between the power of the incentives in the High-Choice treatment (D = 0.452, p = 0.000) but no difference in the High-High treatment (D = 0.452, p = 0.735).

Table 3 reports the results of a series of ordinary least squares (OLS) regressions examining the effect of the incentives and, in general, the determinants of output in the working stage of the experiment for the mission sector. The outcome variable is the number of matrices solved in the second stage for the mission sector.

Independent from the specification, the number of matrices solved in the first stage (the productivity measure) is strongly related to the number solved in the second stage. The treatment effects for High-High are small but become significant once interacted with the productivity, suggesting that as productivity increases, so does the number of matrices solved in the mission sector under the performance-based incentive compared to the low-powered one. In other words, as ability increases, the increase in the number of matrices solved is larger in High-High compared to High-Low. This effect is confirmed by Figure A.3. It is mostly driven by higher-ability subjects working harder under the high-powered incentive compared to the low-powered one.

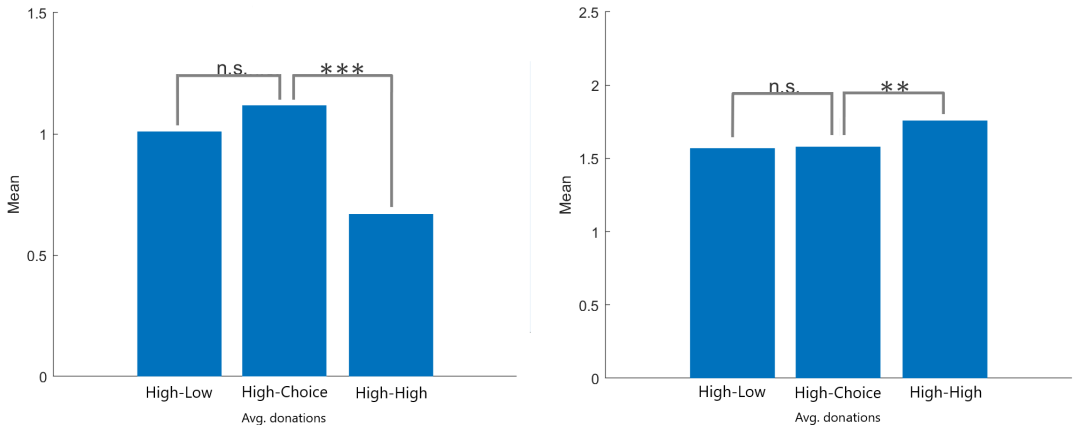
As mentioned by Lazear (2000), payment schemes entail both an incentive and a sorting dimension. The incentive effect motivates current workers, especially more productive ones, to work more. The sorting effect ensures that they choose the type of incentive in which payment depends on performance. Between the prosocial preferences, only positive reciprocity is positively and weakly related to the outcome variable.

These findings suggest that high-powered incentives increase the number of matrices solved in the mission-oriented sector. This is in line with one of the main arguments favouring using high-powered incentives in the prosocial sector, which motivates the use of such incentives with a perspective of an increase in productivity in the public/mission sector.

4.3.1 Charitable donations

In terms of charitable donations, Figure 6 shows the average charitable donation by treatment, both ex ante and ex post. The difference is that the ex ante graphs capture the average donation per participant (this includes those who have selected the profit sector and thus donated £0). In contrast, the ex post graphs capture the average donation per participant, conditional on sorting into the mission sector.

FIGURE 6. Average ex ante and ex post charitable donations



(A) Ex ante donations (B) Ex post donations

Notes: The stars represent the significance levels of the Wilcoxon ranksum tests reported in the text. * p < 0.1, ** p < 0.05, *** p < 0.01

In Figure 6a, the High-Choice treatment reached the highest average ex ante donation to charities. These subjects donated £1.12 (sd: 0.82), those in the High-High treatment donated £0.68 (sd: 0.91), whereas those in the High-Low treatment donated £1 (sd: 0.81). Resulting in a statistically significant difference between the last two treatments (Wilcoxon rank-sum test, $z = 2.952$, $p = 0.003$, $N = 247$) but not between the first two (Wilcoxon rank-sum test, $z = -0.916$, $p = 0.360$, $N = 246$). If we focus on Figure 6b, we can see that the ranking of treatment in terms of donations is almost inverted. After subjects make their choice, the treatment with the highest average charitable donation becomes High-High with £1.77 (sd: 0.51) compared with £1.57 (sd: 0.39) in High-Low and £1.58 (sd: 0.47) donated in High-Choice. Again, the difference between the first two treatments is not statistically significant (Wilcoxon rank-sum test, $z = 0.099$, $p = 0.921$, $N = 166$), whereas the difference between the second and the third treatment it is, but in the opposite direction compared to the ex-ante measure (Wilcoxon rank-sum test, $z = -1.891$, $p = 0.048$, $N = 136$).

There is a one-to-one relationship between the number of matrices solved in the mission sector and the amount of money donated to the chosen charity. It follows that ex post the treatment with the highest average donation is also the treatment with the highest average number of matrices solved. In sum, depending on whether we focus on the ex ante or ex post donations, the treatment with the highest donation is either High-Choice or High-High. To illustrate the difference, consider a case in which a charity needs to hire street fundraisers.

Allowing applicants to choose the type of salary they prefer will result in more applicants. Still, by paying the fundraisers a performance-based salary, the charity would raise more funds per fundraiser at the end of the day. One such payment scheme will involve paying workers a fixed salary plus a bonus if they exceed a certain number of donations collected.

Assuming that the charity's objective is to maximise the amount of donations, to choose which incentive scheme is more suited to this end, it is necessary to know the costs related to the adoption. For example, it could be that the bonus payments made under the high-powered incentive exceed the profits gained by the additional productivity. Also, there could be other limitations, such as hiring constraints to consider, which would make the use of one incentive more profitable than the other. Finally, given that the low-powered incentive does not perform badly, especially ex-ante, it could also be tempting to keep the salary fixed. However, such incentives attract a significant share of low-productivity individuals to the mission sector, which could negatively impact service provision in real life.

5 Discussion and conclusions

The use of high-powered incentives has often been proposed as a way to increase productivity in the mission sector. Performance-based payments entail both an incentive and a sorting dimension. The first makes the current workers work more, whereas the second affects future workers through self-selection; i.e. more productive workers will apply for the position. Previous research suggests that monetary incentives have the potential to crowd out intrinsic motivation. I combine the literature on incentives with the one on crowding out and explore, using a laboratory experiment, whether high-powered incentives affect self-selection into the mission sector, focusing primarily on productivity and altruism. In other words, I analyse how individual characteristics interact with the power of the incentives to determine self-selection into different sectors. I also study the effect of high-powered incentives on productivity in both sectors.

When subjects are faced with a high-powered incentive in the profit sector and a low-powered incentive in the mission sector (the standard setting), altruism is a significant predictor of the prosocial sector choice. Still, I find a productivity gap between the profit and mission sector. When subjects are asked to choose the type of incentive in the prosocial sector, productivity is a significant determinant of choice between incentives. In contrast,

altruism is a significant predictor of choice between sectors. In other words, as productivity increases, subjects are more likely to choose the high-powered incentive, whereas as altruism increases, they are more likely to select the mission sector. In this context, those selecting the high-powered incentive in the mission sector are more productive than those choosing the low-powered one and more altruistic than those choosing the profit sector, which is the best possible combination. When the entire mission sector switches to high-powered incentives, when subjects are faced with a high-powered incentive in both sectors, I observe that productivity increases, but average altruism decreases and stops being a significant predictor of the mission-sector choice. This could be due to a weakening of the relationship between altruism and the mission sector's choice, perhaps due to a shift in the focus from a prosocial to a productivity-oriented goal. Although, as trust decreases when incentives in the social sector are performance-based, it could also be driven by a decrease in perceived trustworthiness. These findings are particularly relevant for all the mission-oriented jobs that entail a strong prosocial component as postings, especially in rural areas, require subjects to exert effort that is not always strictly observable and, therefore, rely on prosocial motivation to reach those that are most in need. Finally, I find that under the high-powered incentives fewer subjects choose to work for the mission sector resulting in a lower level of charity donations overall, despite a better performance on average. This is particularly important, given that one of the main reasons why this type of incentives are adopted is to increase productivity, but what my results show is that they might also have indirect effects which impact profitability.

In developing countries, selecting figures such as Community Health Promoters that are lower on the altruism scale might make a difference in whether the basic health services they promote reach the poorest households. For example, they could mainly focus on selling their products to more affluent households, thus maximising their profit, but in doing so, they would not realise the full extent of the program's benefits. The previous example illustrates why it is important to select altruistic and productive workers to achieve the maximum potential in which case, the best option is to let the subject choose the power of the incentive.

Taken together, these findings suggest that it is possible to screen for both altruism and productivity in the mission sector with high-powered incentives, as long as they are used alongside low-powered ones. Alternatively, by considering the entire labour market, as long as there are organisations paying low-powered incentives, which are the most common payment

type, other mission-driven organisations, such as NGOs, would benefit from choosing high-powered incentives, as it would allow them to attract both productive and prosocial workers. The reason is that given their high productivity; these subjects always have the alternative option to work in the profit sector, hence their choice to work in the mission sector can only be based on prosociality. However, it is important to note that as the mission sector switches entirely to high-powered incentives, this type of incentive ceases to be strictly beneficial. Therefore, the number of firms changing their payment scheme to a high-powered one influences how the subjects distribute among them and ultimately determines which type of applicants they will attract. Potentially, the more firms adopt a high-powered incentive, the lower gets the benefit of adopting such an incentive for each individual firm.

Ultimately, the optimal incentive scheme depends on a series of factors. The best incentive type may vary depending on the role, more or less in contact with the public, the measurability of output and whether the role entails more than one task. However, the results presented in this work suggest that care is needed when choosing the incentive scheme. There might be a trade-off between altruism and productivity, which is important for mission-driven organisations to consider. The solution could be to choose, where possible, different types of incentives depending on the specific role within the mission organisation. For example, it would make sense to hire a hospital manager and pay him based on performance, but it might not be a good idea to pay a nurse based on the number of patients she looks after.

While most previous research has focused either on the consequences of an increase in fixed payments or on the incentive effect related to high-powered payments, this work shows that they also convey important implications in terms of sorting. However, this study did not explore the exact mechanism by which prosocial motivation is crowded out in the case where all firms switch to the high-powered incentive. Future research should examine the effects of high-powered incentives in more detail by isolating the different channels by which they can affect prosocial motivation in the mission sector. Finally, it could also be interesting to bring the same research questions out of the lab and test whether high-powered incentives impact the prospective applicant pool's average prosociality for a mission-oriented position (e.g. for an NGO) and how they affect service provision in a field setting.

A - Supplementary Tables and Figures

A.1 Tables

	(1) High-Low		(2) High-High		(3) High-Choice		(4) Overall		(5) Difference
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	p-value
Demographics									
Female*	0.48	(0.50)	0.63	(0.48)	0.68	(0.47)	0.60	(0.49)	0.003
Age	21.70	(3.06)	22.26	(4.04)	21.63	(3.52)	21.86	(3.56)	0.457
Math	6.98	(2.39)	6.62	(2.38)	6.98	(2.22)	6.85	(2.33)	0.303
Income	317.95	(239.81)	333.15	(294.13)	316.87	(336.83)	322.52	(292.16)	0.700
Main Variables									
N. solved 1	7.36	(2.89)	7.61	(2.70)	6.90	(2.84)	7.29	(2.81)	0.119
N. solved 2	16.75	(4.35)	17.47	(5.10)	16.59	(5.12)	16.94	(4.87)	0.353
Profit*	0.36	(0.48)	0.62	(0.49)	0.29	(0.46)	0.42	(0.49)	0.000
Mission_{Low}*	0.64	(0.48)	0.00	(0.00)	0.50	(0.50)	0.38	(0.49)	0.000
Mission_{High}*	0.00	(0.00)	0.38	(0.49)	0.21	(0.41)	0.20	(0.40)	0.000
Risk	9.46	(3.53)	8.44	(3.07)	8.65	(3.21)	8.84	(3.29)	0.046
Patience	14.00	(8.14)	13.96	(8.19)	13.64	(7.83)	13.87	(8.03)	0.889
P. reciprocity	9.03	(2.74)	9.48	(2.88)	9.47	(2.73)	9.33	(2.78)	0.304
N. reciprocity	4.14	(1.81)	4.42	(1.81)	4.09	(2.04)	4.22	(1.89)	0.291
Trust	5.12	(2.27)	4.89	(2.24)	4.93	(2.37)	4.97	(2.29)	0.831
Altruism	72.56	(62.38)	60.32	(59.63)	60.30	(52.18)	64.38	(58.25)	0.242
Self-control	38.27	(8.38)	38.94	(8.27)	39.95	(8.48)	39.09	(8.39)	0.427
Social desirability	10.00	(2.51)	10.09	(2.78)	9.87	(2.73)	9.99	(2.67)	0.698
N	122		125		124		371		

Variables are binary when asterisked. The reported p-values are obtained from a χ^2 test for binary, and from a Kruskal-Wallis test for the rest of the variables. Age is measured in years, math ability is self-reported on a 10-point scale, where 0 indicates not good at all and 10 very good. Income is measured in pounds.

TABLE A.1. Summary statistics

A.2 Figures

TABLE A.2. Correlation between independent variables

	Productivity 1	Altruism	Female	Risk	Patience	P. reciprocity	N. reciprocity	Trust	Self-control	Social desirability
Productivity 1	1.000									
Altruism	-0.006	1.000								
Female	-0.206***	0.115**	1.000							
Risk	0.029	0.128**	-0.162***	1.000						
Patience	0.044	-0.114**	-0.046	-0.112**	1.000					
P. reciprocity	-0.004	0.197***	0.003	0.075	0.053	1.000				
N. reciprocity	0.081	-0.130**	-0.068	-0.003	-0.023	-0.088*	1.000			
Trust	0.033	0.054	0.004	0.046	0.079	0.062	-0.136***	1.000		
Self-control	-0.138***	0.194***	-0.081	0.057	-0.023	0.090*	-0.223***	0.022	1.000	
Social desirability	-0.070	0.157***	-0.109**	0.156***	-0.013	0.123**	-0.137***	0.170***	0.411***	1.000

Pairwise correlations between the independent variables.

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

	High-Low		
	$Altruism_L$	$Altruism_H$	Total
$Productivity_L$	17	29	46
	21.79	37.17	58.97
$Productivity_H$	16	16	32
	20.51	20.51	41.50
Total	33	45	78
	42.30	57.69	100.00

	High-High		
	$Altruism_L$	$Altruism_H$	Total
$Productivity_L$	12	12	24
	25.00	25.00	50.00
$Productivity_H$	13	11	24
	27.08	22.92	50.00
Total	25	23	48
	52.08	47.92	100.00

	High-Choice		
	$Altruism_L$	$Altruism_H$	Total
$Productivity_L$	27	32	59
	30.68	36.36	67.04
$Productivity_H$	11	18	24
	12.50	20.45	27.27
Total	38	50	88
	43.18	56.82	100.00

TABLE A.3. Composition of the mission sector across treatments

	High-Low		
	Male	*Female*	Total
Profit	29	15	44
	45.31	25.86	36.07
Mission	35	43	78
	54.69	74.14	63.93
Total	64	58	122
	100.00	100.00	100.00

	High-Choice		
	Male	*Female*	Total
Profit	14	22	36
	35.00	26.19	29.03
Mission	26	62	88
	65.00	73.81	70.97
Total	40	84	124
	100.00	100.00	100.00

	High-High		
	Male	*Female*	Total
Profit	31	46	77
	67.39	58.23	61.60
Mission	15	33	48
	32.61	41.77	38.40
Total	46	79	125
	100.00	100.00	100.00

TABLE A.4. Distribution of males and females across treatments

	(1) $Profit$	(2) $Mission_L$	(3) $Mission_H$
Productivity	0.171***	-0.320***	0.149***
	(0.055)	(0.076)	(0.042)
Altruism	-0.136**	0.081	0.055
	(0.068)	(0.066)	(0.042)
Risk	0.042	-0.088	0.046
	(0.049)	(0.058)	(0.046)
Patience	-0.057	0.010	0.047
	(0.054)	(0.058)	(0.038)
P. reciprocity	0.013	-0.030	0.017
	(0.047)	(0.052)	(0.043)
N. reciprocity	0.034	-0.012	-0.022
	(0.048)	(0.058)	(0.041)
Trust	0.026	0.020	-0.046
	(0.051)	(0.053)	(0.036)
Self-control	-0.047	-0.031	0.078*
	(0.046)	(0.054)	(0.043)
Social desirability	0.137***	-0.094	-0.042
	(0.053)	(0.059)	(0.046)
Controls	Yes	Yes	Yes
Observations	124	124	124

Notes: Multinomial logit estimates. Marginal effects (evaluated at the mean of the independent variables) reported; robust standard errors between parenthesis. The outcome variable takes value one if the subject chose the mission sector. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

TABLE A.5. Determinants of the choice of sector and incentive type for the High-Choice treatment

	(1) High-Low	(2) High-Low	(3) High-Low
Productivity	-0.105**	-0.097**	-0.101**
	(0.046)	(0.046)	(0.048)
Altruism	0.072*	0.112**	0.145**
	(0.043)	(0.055)	(0.057)
Risk		-0.029	-0.041
		(0.045)	(0.046)
Patience		0.057	0.041
		(0.049)	(0.048)
P. reciprocity		-0.051	-0.001
		(0.052)	(0.055)
N. reciprocity		-0.097	-0.136**
		(0.061)	(0.064)
Trust		-0.083	-0.067
		(0.052)	(0.051)
Self control			-0.054
			(0.058)
Social desirability			-0.111**
			(0.054)
Controls	No	Yes	Yes
Pseudo R-squared	0.053	0.063	0.206
Observations	122	122	122

Notes: Logit estimates. Marginal effects (evaluated at the mean of independent variables) reported; robust standard errors between parenthesis. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

TABLE A.6. Determinants of the choice of the mission sector for the High-Low treatment

	(1) High-Choice	(2) High-Choice	(3) High-Choice
Productivity	-0.105***	-0.103**	-0.123***
	(0.040)	(0.040)	(0.040)
Altruism	0.092*	0.110*	0.132**
	(0.047)	(0.057)	(0.063)
Risk		-0.051	-0.030
		(0.042)	(0.043)
Patience		0.044	0.061
		(0.043)	(0.047)
P. reciprocity		0.005	-0.001
		(0.044)	(0.041)
N. reciprocity		-0.022	-0.026
		(0.038)	(0.039)
Trust		-0.052	-0.033
		(0.042)	(0.044)
Self control			0.050
			(0.040)
Social desirability			-0.121***
			(0.045)
Controls	No	Yes	Yes
Pseudo R-squared	0.053	0.063	0.144
Observations	124	124	124

Logit estimates. Marginal effects (evaluated at the mean of independent variables) reported; robust standard errors between parenthesis. The outcome variable takes value one if the subject chose the mission sector. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

TABLE A.7. Determinants of the choice of the mission sector for the High-Choice treatment

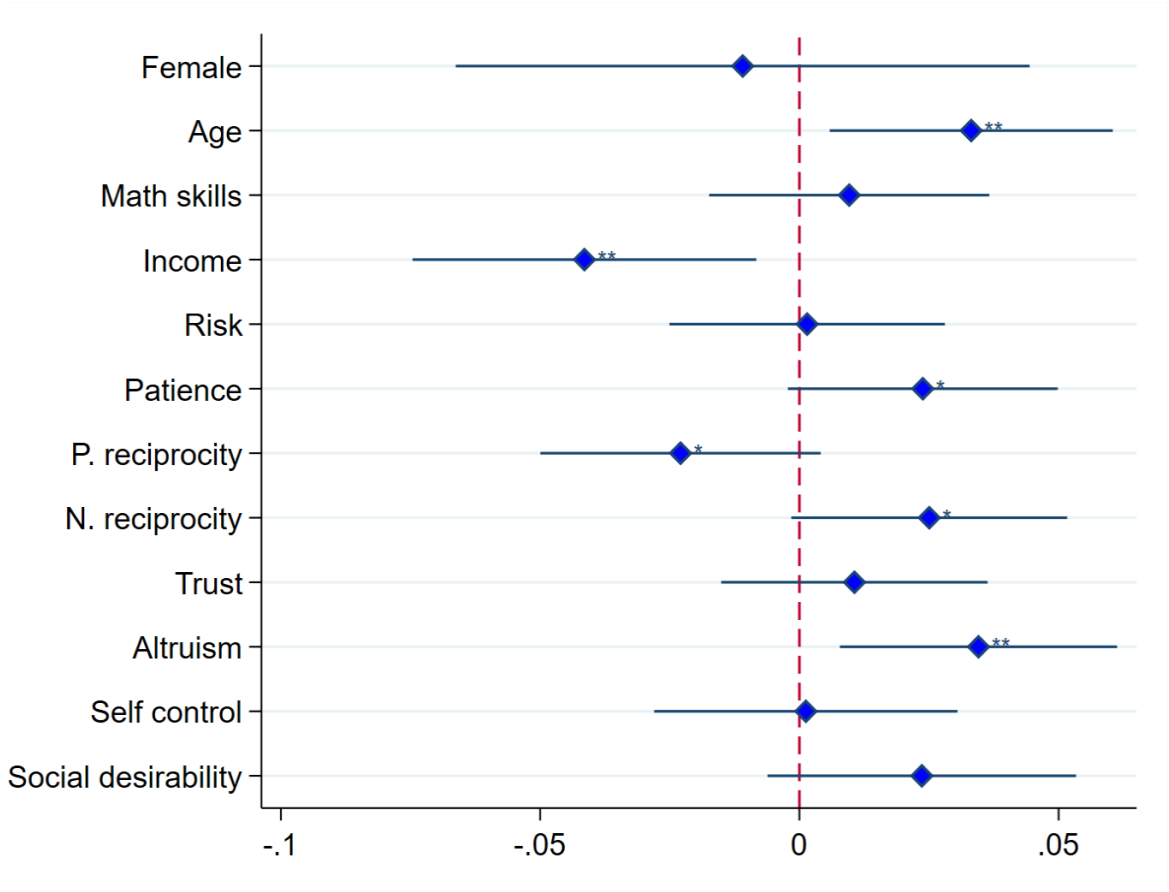
	(1) High-High	(2) High-High	(3) High-High
Productivity	-0.000	0.006	0.013
	(0.046)	(0.047)	(0.046)
Altruism	0.055	0.042	0.039
	(0.044)	(0.044)	(0.046)
Risk		-0.006	-0.018
		(0.050)	(0.051)
Patience		-0.040	-0.040
		(0.044)	(0.044)
P. reciprocity		0.067	0.068
		(0.050)	(0.051)
N. reciprocity		-0.092*	-0.092*
		(0.047)	(0.049)
Trust		-0.072	-0.090**
		(0.044)	(0.046)
Self control			-0.018
			(0.061)
Social desirability			0.072
			(0.052)
Controls	No	Yes	Yes
Pseudo R-squared	0.010	0.012	0.081
Observations	125	125	125

Notes: Logit estimates. Marginal effects (evaluated at the mean of independent variables) reported; robust standard errors between parenthesis. The outcome variable takes value one if the subject chose the mission sector. All independent variables are standardized to mean zero and standard deviation one. The controls include: Age, Female and Income.

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

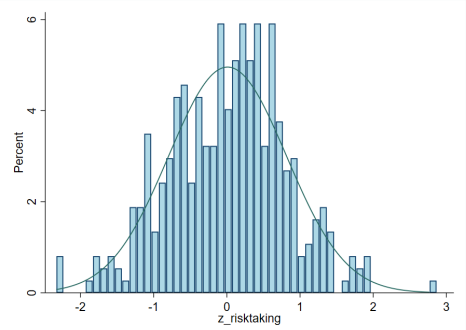
TABLE A.8. Determinants of the choice of the mission sector for the High-High treatment

FIGURE A.1. Coefficient plot examining the determinants behind the subject's choice to inscribe to the second stage of the experiment

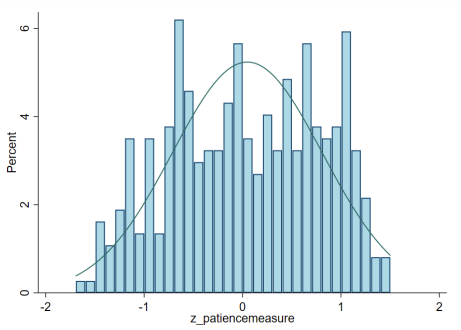


Notes: Marginal effects (evaluated at the mean of independent variables) and 95% confidence intervals for a logit regression. The outcome variable takes value one if the subject participated in the first stage of the experiment and inscribed to the second one, and zero otherwise. All independent variables are standardized to mean zero and standard deviation one.

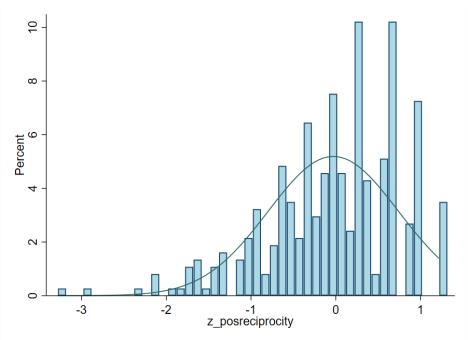
FIGURE A.2. Distribution of standardised preference measures



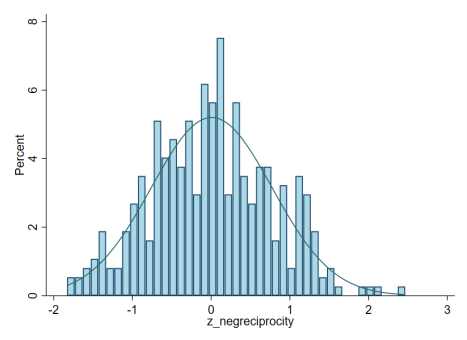
(A) Risk



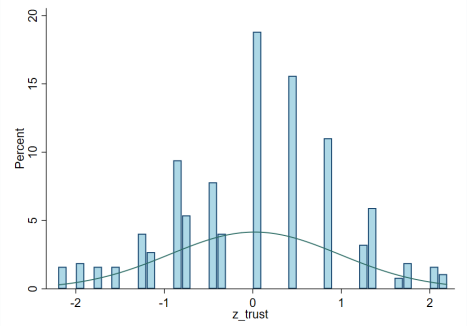
(B) Patience



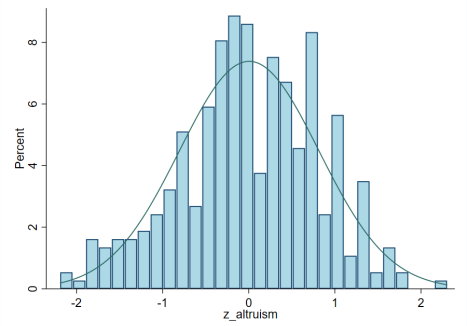
(C) Positive reciprocity



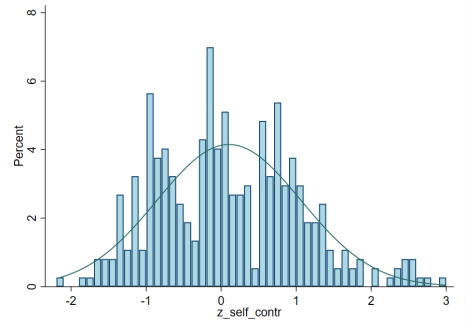
(D) Negative reciprocity



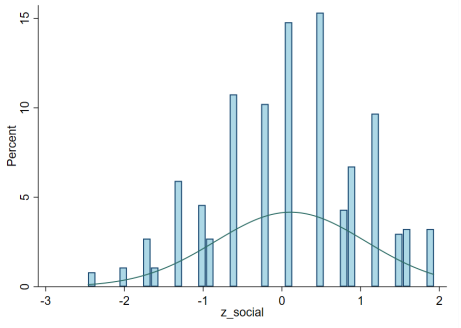
(E) Trust



(F) Altruism

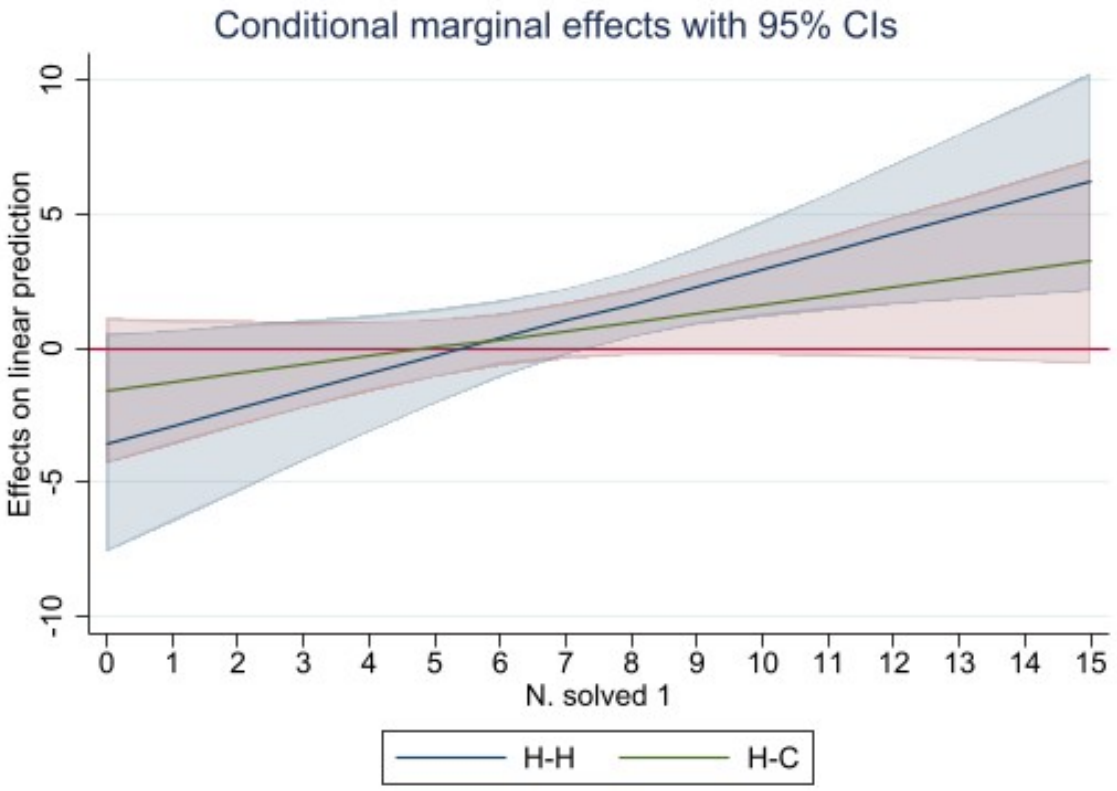


(G) Self-control



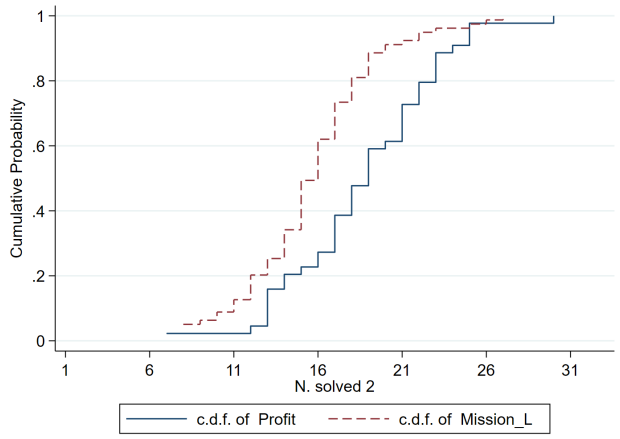
(H) Social desirability

FIGURE A.3. Marginal effects examining the interaction between treatments and productivity

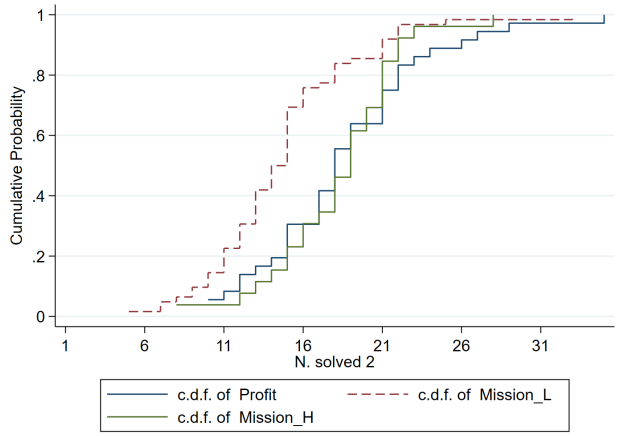


Notes: Marginal effects (evaluated at the mean of independent variables) and 95% confidence intervals for an OLS regression. The outcome variable is the number of matrices solved in the second stage.

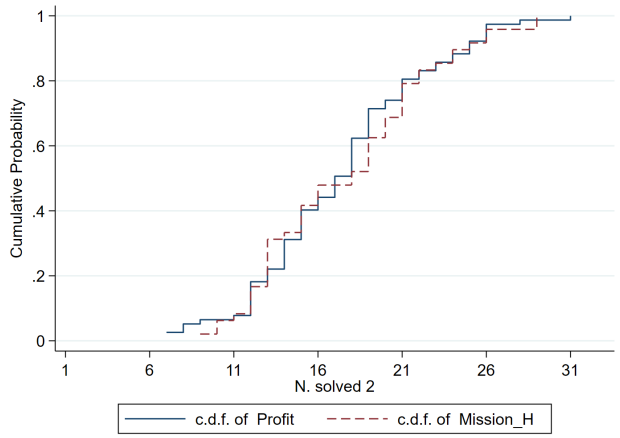
FIGURE A.4. Cumulative density function of individual output in the working stage of the experiment by sector and treatment



(A) High-Low



(B) High-Choice



(C) High-High

CHAPTER III

Choosing Competition on Behalf of Someone Else*

*Co-authored with Nina Serdarevic and Helena Fornwagner.

Abstract

We extend the existing literature on gender differences in competitive behaviour by investigating tournament entry choices when a principal decides for an agent. In a laboratory experiment, we randomly assign subjects the role of either principal or agent. The principal decides whether the agent performs a real-effort task under piece-rate or tournament incentives. When deciding, the principal is informed about the agent's previous performance, age and residency. Between treatments, we vary whether the principal knows the agent's gender. In a baseline treatment, we replicate the standard setting in which subjects decide for themselves whether to compete. Our main findings are, first, that there is no gender gap in tournament entry when principals decide for agents, as opposed to the baseline treatment. Second, the gender gap closes because more women are made to compete by principals. Third, while there is no gender gap in either of the principal treatments, revealing the agent's gender is associated with higher overall tournament entry rates. Exploratory analyses of principals' choice determinants reveal a positive effect of preferences to take risks, competitiveness and confidence in agents' performances on making agents compete. Besides, we find no difference in how principals evaluate male and female agents' performances. Finally, we test the efficiency of principals' competition choices and show that they lead to fewer payoff-maximising outcomes than when subjects decide for themselves. Additionally, overall tournament performances and winners' performances are lower when agents are made to compete, but this effect is not robust to controlling for agents' previous performances.

JEL Classification: J16, C91, D01

Keywords: competitiveness, gender differences, decision-making for others, laboratory experiment

1 Introduction

Despite important advances over the past century, differences in labour market outcomes between women and men are persistent (see, e.g. European Commission, 2019). Possible explanations include discrimination (Black & Strahan, 2001), anticipated discrimination (Charness, Cobo-Reyes, Meraglia, & Sánchez, 2020; Fisk & Overton, 2019), gender differences in preferences regarding the type of employment, rank or position (Clain, 2000) or difficulty balancing family and career (Mason, Wolfinger, & Goulden, 2013). Another explanation focuses on gender differences in preferences for competition, suggesting that women tend to shy away from competitive environments (see, e.g. Niederle & Vesterlund, 2007). This tendency is particularly problematic if payment schemes involving competition also pay higher wages, and it may partly explain the gender pay gap (Blau & Kahn, 2017). As such, understanding the causes and consequences of the gender differences in preferences to compete has attracted attention in recent years and sparked a vast follow-up literature (for a survey, see Niederle, 2017).

To date, research on competitiveness has examined how individuals make competitive choices when deciding for themselves. However, many important decisions are not entirely up to an individual but are made by others (Füllbrunn, Luhan, & Sanfeycd, 2020).[1] This raises the question of whether the gender differences in choosing to compete extend to principals, either because they treat male and female agents differently or because male and female principals make different choices. Although the literature on gender and competitiveness has grown considerably over the past decade, these questions remain largely unexplored. In this paper, we ask whether there is a gender difference in tournament entry when the decision to compete is not made by an individual (i.e. an agent) but by someone else who faces the same monetary consequences of the tournament entry choice (i.e. a principal). Moreover, we examine how the entry rates of men and women differ when deciding for themselves as opposed to being made to compete and finally, we causally investigate the

[1] Examples of such situations in Füllbrunn et al. (2020) include parents deciding for children, CEOs deciding in the interests of companies, or, more generally, trustees deciding for beneficiaries.

role of gender information when deciding for others. Besides, we explore the role of other determinants that may underlie the choice to make others compete and, importantly, the efficiency consequences related to principals deciding for agents.

We conduct a laboratory experiment employing a modified version of the Niederle and Vesterlund (2007) design. Three treatments – *GenderInfo*, *NoGenderInfo* and *OwnDecision* – are implemented using a between-subjects design. In *GenderInfo* and *NoGenderInfo*, we randomly assign subjects the role of either principal or agent. Each principal is matched with one agent. The agents perform a real-effort task, first under a piece-rate (Stage 1) and then a tournament (Stage 2) payment scheme. Then, the principal chooses which scheme to apply to their agent's future performance in the real-effort task (Stage 3). In *GenderInfo*, the principals are informed of their agent's absolute performance in Stage 1 and Stage 2 as well as their gender, age and UK residency. In *NoGenderInfo*, principals receive the same information except the agent's gender. In both treatments, if Stage 3 is randomly chosen for payment, the principal and agent are paid according to the agent's performance in Stage 3 such that the monetary incentives of principals and agents are aligned. We compare the principals' tournament entry choices and the agents' performances to *OwnDecision*, in which subjects choose for themselves under which incentive scheme to work in Stage 3.

In line with the existing literature, we find a significant gender gap in competitiveness (GGC) when subjects decide for themselves; in particular, men are 1.6 times more likely than women to compete. However, when a principal chooses on behalf of an agent, there is no GGC. Comparing choices in *GenderInfo* to *OwnDecision* suggests that the GGC closes as a consequence of significantly more women being made to compete when principals decide. Furthermore, a comparison of *GenderInfo* and *NoGenderInfo* allows us to discern the extent to which information about the agent's gender affects the principal's choice. While there is no GGC in either of the principal treatments, revealing the agent's gender results in higher overall tournament entry rates in *GenderInfo* compared to *NoGenderInfo*. Although we did not hypothesise this result ex-ante, we discuss some potential reasons for why this may be the case.

An exploratory analysis of the determinants behind the principal's decision shows that the

principal's gender and the interaction of the principal's and the agent's genders do not affect their choice of payment scheme. However, more competitive and risk-seeking principals are more likely to choose competition for their agents; hence, the principals' preferences matter for their choices. Whereas we observe a gender gap in confidence among subjects performing the real-effort task in *OwnDecision* and *GenderInfo*, principals do not evaluate the relative performance of female agents differently from males in *GenderInfo*. Relatedly, agents who are perceived to perform better in the tournament are more likely to be made to compete by their principals, suggesting that the primary motivation of the principals is to maximise earnings by sending who they perceive to be the best agents to the tournament. Finally, in terms of efficiency, our results suggest that closing the GGC in our setting comes at the cost of principals making less payoff-maximising decisions. Moreover, the average tournament performance and the performance of winning subjects are lower when agents are made to compete by principals. Still, this effect is not robust to controlling for subjects' general abilities in the real-effort task.

Choosing on behalf of others is an integral part of life and professional environments, and gender may be an essential factor in shaping the perceived suitability of individuals in competitive environments. Using a controlled laboratory environment, our paper makes a two-fold contribution to the literature concerned with these questions: First, we show that the commonly documented GGC does not carry over to situations in which a principal decides for an agent. Second, we examine the role of the agent's gender when principals choose while controlling for factors such as performance, preferences and beliefs that may underlie the decision to make the agent compete; indeed, in many naturally occurring situations, principals have more accurate information about agent's preferences and abilities than in our laboratory setting. However, estimating the causal impact of gender information with observational data may be challenging or even impossible due to several confounding factors. One such factor is that genders of the principal and the agent cannot be exogenously manipulated, as in an experiment, because these are typically known. Moreover, selection by principals and agents may cause unobservable characteristics to correlate with choices, performance and other important outcome variables.

The rest of the paper is structured as follows: **Section 2** summarises the related literature, **Section 3** outlines the experimental design, **Section 4** provides the data analysis and results and **Section 5** concludes the paper.

2 Related literature

2.1 Willingness to compete

Over the last decades, a vast literature documenting gender differences in competitiveness has accumulated. This literature generally suggests that fewer women enter competitive environments than men when deciding for themselves (see, e.g. Balafoutas & Sutter, 2019; Saccardo, Pietrasz, & Gneezy, 2018; Niederle, 2017; Almås, Cappelen, Salvanes, Sørensen, & Tungodden, 2016b; Sutter & Glätzle-Rützler, 2015; Datta Gupta, Poulsen, & Villeval, 2013; Niederle & Vesterlund, 2007). This behavioural pattern has been replicated in various contexts and is robust to different specifications. Nonetheless, some exceptions include matriarchal societies in which women and men are equally competitive (Andersen, Ertac, Gneezy, List, & Maximiano, 2013) or women are more competitive than men (Gneezy, Leonard, & List, 2009), and girls' schools, where female students are more competitive compared to those enrolled in mixed-sex schools (Booth & Nolen, 2012). Furthermore, the GGC disappears when women compete against themselves rather than against someone else (Apicella, Demiral, & Mollerstrom, 2017) or when they compete as part of a team (Dargnies, 2012).

A person's competitiveness, as measured in the economic lab, is important as it correlates with various choices and characteristics relevant for education and labour market outcomes. For example, less competitive individuals are less likely to apply for jobs in which their salaries are tied to how well they perform compared to others (Flory, Leibbrandt, & List, 2015). Moreover, those who are more competitive in lab experiments have been found to be more likely to choose competitive educational programs (e.g. Buser, Niederle, & Oosterbeek, 2020; Reuben, Wiswall, & Zafar, 2017; Almås, Cappelen, Salvanes, Sørensen, & Tungodden, 2016a; Buser, Niederle, & Oosterbeek, 2014), have a higher income (e.g. Buser, Geijtenbeek,

& Plug, 2018; Kamas & Preston, 2015; Reuben, Sapienza, & Zingales, 2015) and become entrepreneurs (Berge, Bjorvatn, Pires, & Tungodden, 2015).

Given the implications of these findings, a body of research has looked for ways to close the GGC. Some studies change the institutional environment to resemble different affirmative-action policies to obtain gender balance in competitive environments. Examples include quotas (Baldiga & Coffman, 2018; Leibbrandt et al., 2018; Niederle et al., 2013), preferential treatment (Balafoutas & Sutter, 2012) and allowing the prize to benefit one's offspring (Cassar et al., 2016). While many interventions have been shown to narrow the GGC, there is growing evidence that they tend to backfire on those favoured (Fallucchi & Quercia, 2018; Leibbrandt et al., 2018).

To circumvent these drawbacks and offer more easily implementable interventions, another stream of the literature has documented that priming subjects with power (Balafoutas, Fornwagner, & Sutter, 2018) or a professional, work-related identity (Cadsby, Servátka, & Song, 2013) encourages women to enter competitions more often. Also, giving feedback about relative performance (Wozniak, Harbaugh, & Mayr, 2016) and the earning implications related to competition avoidance (Kessel, Mollerstrom, & van Veldhuizen, 2019) successfully increases women's entry rates into tournaments. In addition, advice from more experienced people improves the entry of strong-performing women into competitions, making them more confident about their performance (Brandts, Groenert, & Rott, 2015).

2.2 Taking decisions for others

Whereas the majority of the existing literature on competitiveness focuses on individual decisions to enter a tournament, our paper adds to the more recent literature on making decisions on behalf of others in various domains (Füllbrunn et al., 2020). When it comes to choosing for others in competitive settings, Tungodden (2019) shows that parents choose more competition for boys than for girls. Nonetheless, considering that a child's gender is always known to their parents, it becomes challenging to tease out the role of gender in choosing for others. Our experimental design allows us to causally investigate whether the GGC extends to situations in which principals decide.

A study conducted independently and in parallel to ours also examines whether the GGC persists in the context of decision making for others and leads to similar results (Ifcher & Zarghamee, 2020), although building on somewhat different grounds. Our study differs in several aspects: First, while principals in Ifcher and Zarghamee (2020) are always informed about the agent's gender, we vary whether the gender information is provided between treatments. This allows us to causally examine the impact of gender information on making others compete. Furthermore, Ifcher and Zarghamee (2020) have subjects decide both for themselves and for someone else whether they should enter the tournament. In our case, principals decide for someone else and also report hypothetical choices for themselves. The reason why we kept these choices as hypothetical choices is to avoid anchoring and spillovers to the actual decision to make others compete, which is our main focus. Finally, Ifcher and Zarghamee (2020) vary the principals' financial incentives in the game. In one case, they have no financial incentive, and in another, they receive 10% of the gains from their agent's payment. In our experiment, principals' incentives are always entirely in line with their agent's compensation.

Despite these differences, the results are very similar; there is no GGC when deciding for others. However, in our case, the gender gap in tournament entry closes because more women are being made to compete. In their case, it closes due to fewer men being sent into the competition. The similarities of results are reassuring, especially considering the recent prominence given to replication studies in the discipline.

In the risk domain, people in positions of authority are often called upon to select courses of action involving risk to themselves and others. So far, studies have shown mixed results. Some papers find that subjects take on more risk when investing money on behalf of others, whereas other papers find increased risk aversion (Eriksen et al., 2020; Friedl, Pondorfer, & Schmidt, 2020). Slovic, Weinstein, and Lichtenstein (1967) show experimentally that women and men, on average, make the same risky choices for others as for themselves. Our paper engages with this strand of the literature, as sending an agent into the tournament is a risk-loaded choice, even if the principals do not have to compete themselves. Moreover, we can control how risk preferences shape the principal's choice for the agent with our

experimental setup.

In our setting, a principal always decides on behalf of an agent, and the agent must follow the principal's decision. Thus, more broadly, our paper also adds to the literature on delegation, which explores *whom* principals appoint to represent them. For example, varying whether the agent's gender is revealed in a dictator game with compulsory delegation, Bottino, García-Muñoz, and Kujal (2016) show that male and female principals are more likely to delegate decisions to male agents over time. For a general summary of the delegation literature, see Erat (2013), Hamman, Loewenstein, and Weber (2010) and Fershtman and Gneezy (2001b).

2.3 Efficiency

Finally, we contribute to the literature investigating efficiency concerns related to narrowing the GGC. On the one hand, allowing principals to choose can have efficiency-enhancing effects by making high-performing individuals enter a competition. Such results have been obtained by affirmative-action policies and more 'soft' interventions, such as advice from third parties (Brandts et al., 2015; Balafoutas & Sutter, 2012). On the other hand, efficiency losses can occur, as, for example, individuals often have a better grasp of their abilities and preferences than those choosing on their behalf. For example, Exley, Niederle, and Vesterlund (2020) find that, when forced to negotiate, women achieve worse outcomes than when they decide whether to negotiate on their own. This study also suggests a paternalistic demand to force more women to enter negotiations than men.[2] Given the motivations that may be at play when choosing for others, it is not clear a priori whether principals in *GenderInfo* and *NoGenderInfo* or decision-makers in *OwnDecision* will make more efficient tournament entry choices.

[2]Related insights are offered by Tungodden (2019) who finds that paternalistic parents are more likely to enrol a daughter than a son into a tournament. For a more general discussion on paternalistic decision-making, see Ambuehl, Bernheim, and Ockenfels (2021).

3 Experimental design and procedures

Six hundred eighty-eight students (343 men and 345 women) from the University of Nottingham participated in our experiment, conducted in March 2020 at the Centre for Decision Research and Experimental Economics (CeDEx).[3,4,5] Subjects were recruited through ORSEE (Greiner, 2015) following the lab's standard recruitment procedures. Upon arrival, participants knew only that the session was scheduled to last one hour and that it was an experiment in economics. On average, it took 45 minutes for the participants to complete the study. The average payment was £10.40 (including a £4 show-up fee).

For comparability purposes, we build on the design by Niederle and Vesterlund (2007) and use the same real-effort task. The task consists of finding the sum of a series of five randomly drawn two-digit numbers within five minutes, without using a calculator.[6] Each series of numbers is shown on a computer screen together with an input box, a submit button and a counter indicating the number of correctly and incorrectly solved addition tasks. Once the participants submit their answers, the next series of numbers appears, and the counter updates, indicating whether the submitted answer was correct or incorrect. Detailed experimental instructions can be found at the end of the dissertation.

Subjects first perform the task under a piece-rate payment scheme (Stage 1) and then in a tournament (Stage 2) against three other subjects. In Stage 3, depending on the treatment, they are either asked to choose between the piece-rate and the tournament payment, or someone else – a randomly matched principal – decides on their behalf.[7] In the final stage, Stage 4, all participants go through an incentivised Holt and Laury (2002) lottery task to measure their individual risk preference. A random draw decides which stage is chosen for payment; each has an equal probability of being selected. At the end, subjects answer a variety of

[3]The study received ethical approval from the ethics committee of the University of Nottingham on 27 January 2020.

[4]Our experiment was pre-registered on aspredicted.org. The title for the pre-registration is 'Her or Him? Choosing on behalf of someone else' (https://aspredicted.org/blind.php?x=pi2ku4 Number #33924).

[5]All data were collected before UK universities were affected by the COVID-19 pandemic in 2020.

[6]The experiment was programmed using the online platform LIONESS (Giamattei et al., 2020).

[7]We used neutral language in the experiment, calling the subjects in *GenderInfo* and *NoGenderInfo* 'Player A' and 'Player B', respectively, instead of 'principal' and 'agent'.

non-incentivised survey questions. Table 1 provides a timeline of the experimental stages.

ROLE	Stage 1	Stage 2	Stage 3		Stage 4
Decision-maker	*Piece-rate* Real-effort task Belief elicitation	*Tournament* Real-effort task Belief elicitation	Own choice	Real-effort task	Lottery
Agent (A)	*Piece-rate* Real-effort task Belief elicitation	*Tournament* Real-effort task Belief elicitation	Belief about P's choice Own preference	P's choice revealed Real-effort task	Lottery
Principal (P)	Trial	Waiting time	Choice for A Belief elicitation Own preference	Implicit Association Test	Lottery

TABLE 1. Main features of the experimental design

We implement a between-subjects experimental design comprised of three treatments: *OwnDecision*, *GenderInfo* and *NoGenderInfo*. *OwnDecision* mimics the standard Niederle and Vesterlund (2007) design, in which subjects decide for themselves whether to enter the tournament in Stage 3. We refer to subjects in the *OwnDecision* treatment as decision-makers. This treatment serves as a benchmark against which we compare the effect of choosing a payment scheme on behalf of someone else. In *GenderInfo* and *NoGenderInfo*, subjects are randomly assigned to the role of principal or agent, and they remain within this role throughout the study. Each principal is randomly matched with an agent. The principal's task is to decide which payment scheme to apply to their agents' performance in Stage 3. Before choosing, they are provided with information on their screen about their agent's age, UK residency and absolute performance in Stage 1 and Stage 2.[8] The only difference between *GenderInfo* and *NoGenderInfo* is whether the principals receive information about their agent's gender in addition to the above information.[9]

[8]To make sure we convey information correctly, we collected individual characteristics of all subjects independently from their role, at the beginning of the experiment.

[9]Previous studies have conveyed gender in different ways – for instance, with the use of avatars (Charness, Cobo-Reyes, Meraglia, & Sánchez, 2020; Bohren, Imas, & Rosenberg, 2019), names (Brock & De Haas, 2019) or photographs (Castillo & Petrie, 2010) signalling a subject's gender. In the delegation setting of Bottino et al. (2016), gender is revealed by showing principals the ID and gender of an agent on the screen. To avoid making gender too salient, we were inspired by the approach of Bornhorst, Ichino, Kirchkamp, Schlag, and Winter (2010), who hid the emphasis on their main variable of interest by providing several individual characteristics in addition to gender.

	Random role			
Treatments	Principal	Agent	Decision-maker	Total
OwnDecision	/	/	144	144
GenderInfo	136	136	/	272
NoGenderInfo	136	136	/	272
Total	272	272	144	688

TABLE 2. Number of participants by treatment and role

Table 2 shows the number of participants by treatment and role. We obtained the same share of male and female agents and decision-makers in all treatments, ensuring that we had gender-balanced groups (among those who performed in Stage 2 and Stage 3).[10] In *GenderInfo* and *NoGenderInfo*, we had 69 pairs consisting of a male principal and a male agent (MM), 66 pairs of a female principal and a female agent (FF), 67 pairs of a female principal and a male agent (FM), and 70 pairs of a male principal and a female agent (MF). Although the gender of the agent cannot affect the principal's decision in *NoGenderInfo*, to make things more comparable, we made sure we had a similar number of same-sex and mixed-sex pairs.

3.1 Agents/Decision-makers

Here, we describe the procedures and incentives of subjects in the role of decision-makers (*OwnDecision*) or agents (*GenderInfo* and *NoGenderInfo*). These subjects go through four stages, as shown in Table 1. Under the piece-rate scheme, they are paid £0.50 for each correct answer. Under the tournament payment scheme, they are randomly matched in gender-balanced groups of four. Only the person with the highest number of correctly solved addition tasks gets paid £2 per correct answer; the rest of the group receives no payment. In the case of a tie, the winner is chosen randomly. Subjects receive information about their group's gender composition but no further details about the other participants. In each stage, subjects receive feedback on their absolute performance, but they are not informed about

[10]When seating the subjects, we first made sure to have gender balance in the agents' groups. Then we filled the remaining seats on the principal's side of the room. Importantly, we alternated the gender-matching between agents and principals across sessions; for some sessions, principals and agents had the same gender, and for others, the opposite.

their relative performance until the end of the experiment.

In Stage 3, if a principal decides to send an agent into the tournament, the agent's performance is evaluated against the performance of their group members in Stage 2. If the agent solves more addition tasks in Stage 3 than their group members did in Stage 2, they get paid £2 for each matrix correctly solved; otherwise, they earn £0 in that stage.[11] This way, the principal's choice cannot be influenced by the expectation of the other principal's decisions. Subjects assigned to *OwnDecision*, the decision-makers, go through the same stages as the agents, have the same monetary incentives and are asked to guess their relative ranks but decide themselves on the payment scheme in Stage 3 (see Table 1).

Agents are asked to guess their ranking within quartiles (top [1] to bottom [4]) compared to the other participants (Stage 1) or other group members (Stage 2) and are paid £0.50 in addition to their final payoff if their guess is correct. Also, before they discover which payment scheme has been chosen on their behalf, they are incentivised to guess which payment their principal is going to choose and asked which one they would have chosen for themselves.[12] This process allows us to control whether the agents' hypothetical choice affects their performance after discovering which incentive type has been chosen for them. Only one of the incentivised guesses is drawn to be paid at the end of the experiment. In the Holt and Laury (2002) task of Stage 4, the subjects are asked to make a series of 10 pairwise decisions between lotteries, each having an equal probability of being chosen for payment.

3.2 Principals

Principals start the experiment with a two-minute, non-incentivised trial round of the real-effort task. Similar to agents, they see a counter indicating the total of correctly and incorrectly solved addition tasks. The trial round allows principals to familiarise themselves with the real-effort task, enabling them to make a more informed decision when choosing for their agent. Notably, to avoid anchoring and a direct comparison between the principals' and their agents' performance, they have less than half of the time assigned to their agents to

[11] In the case of a tie, the winner is randomly determined.

[12] Agents do not know anything about their principal except that there is a randomly drawn participant in the room who will choose a compensation scheme on their behalf.

perform the task. After the trial stage, the principals are informed that they have been paired with a randomly chosen subject (presented to them as Player A) and are asked to select a payment type on the subject's behalf. If they choose the piece-rate, both receive £0.50 for each correct answer provided by the agent. If they select the tournament, both receive £2 per correct answer if the agent's performance in Stage 3 exceeds that of their other group members in Stage 2. Otherwise, they receive no payment. The principal's decision screen contains information about their agent's absolute performance in Stage 1 and Stage 2, as well as their age, UK residency and, in *GenderInfo*, the agent's gender.

After the principals have chosen, they guessed which incentive scheme the agents would have chosen if they had the chance to and how well they think their assigned agent performed in Stage 1 (relative to the other players in the room) and Stage 2 (relative to the other group members). They then complete an implicit-association test (IAT) (Greenwald, McGhee, & Schwartz, 1998) before moving to the Holt and Laury (2002) lottery task.[13] Additionally, at the end of the experiment, the principals are asked to guess which gender solved more tasks on average, first under the piece-rate and then the tournament payment scheme. One guess is randomly chosen at the end of the experiment, paying £0.50 in addition if correct. We also ask the principals which incentive type they would have chosen for themselves.

4 Results

This section is structured as follows. First, we provide an overview of the summary statistics. Second, we focus on the main results of this paper; the tournament entry decisions and whether they differ when subjects choose for themselves in *OwnDecision* compared to when principals decide on their behalf in *GenderInfo* and *NoGenderInfo*. Third, we explore some of the determinants that may underlie the principal's choice, such as their preferences for risk and competitions, other individual characteristics and the role of confidence in their agent's

[13]The implicit-association test consists of rapidly associating concepts with an attribute (in our case, it was family/career as associated with male or female names), which is meant to capture subconscious bias. We do not include this measure in the main analyses because of technical issues causing missing IAT scores for approximately 4.5% of the principals and making it impossible to match the IAT scores to the individual choices.

relative performance. We conclude the analysis by investigating both ex-ante and ex-post efficiency measures across experimental treatments.

4.1 Summary statistics

Table A.1 summarises our subjects' demographics, which we also include as control variables in our analysis. Of the 688 participants, 345 are women (50.15%), 21.83 years old, on average, and 65.41% are UK residents. For conciseness, we do not present the details of all statistical tests in what follows; instead, we focus on comparing the characteristics of principals and decision-makers as they make the tournament entry choices. We also compare the performance of decision-makers and agents in the real-effort task.

Risk preferences (see Figure A.2), as elicited with the Holt and Laury (2002) lottery task, do not differ between treatments (Kruskal–Wallis test, $p = 0.606$), nor between the decision-makers in *OwnDecision* and the principals in *GenderInfo* and *NoGenderInfo* (Wilcoxon rank-sum test, $p = 0.462$). In line with, e.g. Charness and Gneezy (2012), women are on average more risk-averse than men across all treatments (Wilcoxon rank-sum test, $p < 0.05$). Female decision-makers and principals do not differ in their risk preferences (Wilcoxon rank-sum test, $p = 0.489$). The same applies to male decision-makers and principals (Wilcoxon rank-sum test, $p = 0.750$)

In *OwnDecision*, we elicit preferences for competition through subjects' actual choice of payment scheme. In *GenderInfo* and *NoGenderInfo*, we elicit principals' and agents' stated choices for themselves—what they would have chosen if they were the ones deciding and performing. No differences between treatments are found in the choice to compete ($\chi^2(2)$, $p = 0.627$). Furthermore, what decision-makers choose does not differ from principals' hypothetical choices for themselves ($\chi^2(1)$, $p = 0.867$). Despite the random assignment of subjects to roles and treatments, male principals state on average a higher willingness to compete for themselves compared to female principals in *NoGenderInfo* ($\chi^2(1)$, $p < 0.01$) but not in *GenderInfo* ($\chi^2(1)$, $p = 0.730$). The potential implications of this difference for our main results are explored in Section 4.3.

Figure A.2 depicts the cumulative distributions of the agents' and decision-makers' perfor-

mance in Stage 1 and Stage 2. In Stage 1, men perform better than women in *OwnDecision* (Wilcoxon rank-sum test, $p < 0.05$) and *GenderInfo* (Wilcoxon rank-sum test, $p < 0.10$). We do not find any gender difference in terms of Stage 2 performance within each treatment (Wilcoxon rank-sum tests, $p > 0.134$).[14] Agents' performance increases between Stage 1 and Stage 2 in every treatment (signed-rank tests, $p < 0.01$, for each treatment, respectively), possibly due to learning effects or because subjects become more motivated to exert effort when competing. However, the increase in performance does not differ between men and women (Wilcoxon rank-sum test, $p = 0.905$) or between treatments (Kruskal–Wallis test, $p = 0.936$).

4.2 Gender gap in competitiveness

FIGURE 1. Tournament entry by gender and treatment ($N = 416$)

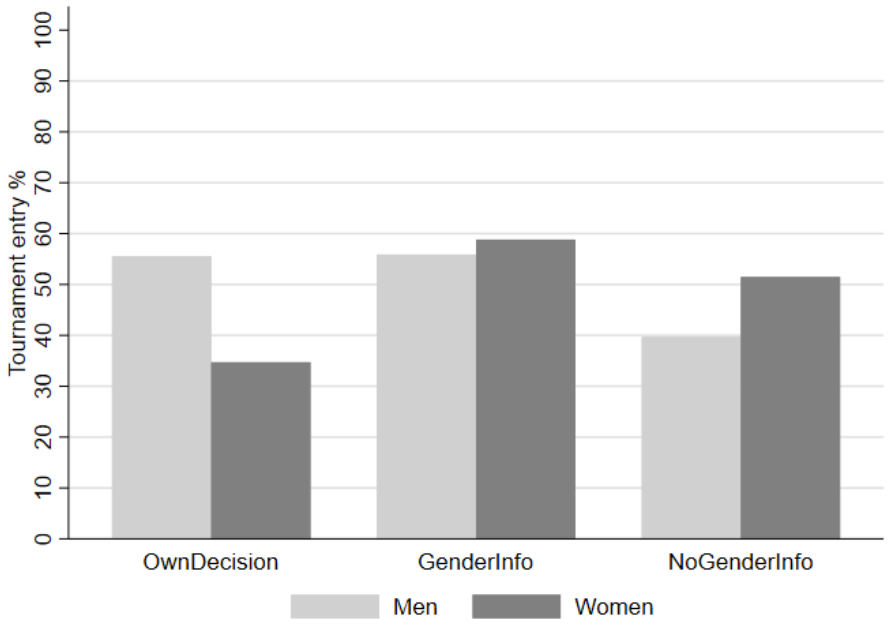


Figure 1 shows the tournament entry of male and female decision-makers and agents by treatment. In *OwnDecision*, 34.72% of women choose to enter the tournament compared to 55.56% of men. The difference is statistically significant ($\chi^2(1)$, $p < 0.01$). Thus, in

[14] Men have sometimes been found to perform better than women under both the piece-rate (Charness, Cobo-Reyes, Meraglia, & Sánchez, 2020) and the competitive payment schemes (Gneezy, Niederle, & Rustichini, 2003).

line with the existing literature showing that women shy away from competitions, we find that men are 1.6 times more likely than women to enter the competition when deciding for themselves. In contrast, principals send male and female agents to the tournament at similar rates in *GenderInfo* (55.88% vs 58.82%; $\chi^2(1)$, $p = 0.729$). Also, no GGC is found in *NoGenderInfo* (39.71% vs 51.47%; $\chi^2(1)$, $p = 0.168$).

	OwnDecision		*GenderInfo*		*NoGenderInfo*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Compete	Compete	Compete	Compete	Compete	Compete
Female	-0.168**	-0.143*	0.030	0.041	0.126	0.087
	(0.081)	(0.077)	(0.087)	(0.081)	(0.086)	(0.084)
Perf. 2 - Perf. 1	-0.001	-0.024	0.005	-0.008	0.017	0.013
	(0.019)	(0.020)	(0.023)	(0.024)	(0.021)	(0.021)
Perf. 2	0.036***	0.024**	0.004	0.006	-0.001	-0.011
	(0.010)	(0.010)	(0.011)	(0.011)	(0.013)	(0.011)
UK resident		-0.198**		-0.046		-0.029
		(0.090)		(0.087)		(0.090)
Age		0.020*		-0.010		-0.007
		(0.011)		(0.014)		(0.012)
Risk		0.027		0.022		0.055**
		(0.020)		(0.028)		(0.023)
Belief 2		-0.100**		-0.171***		-0.209***
		(0.045)		(0.056)		(0.055)
Self.-comp.				0.218***		0.128
				(0.083)		(0.080)
Constant	0.154	0.045	0.517***	0.853**	0.394**	0.759**
	(0.123)	(0.365)	(0.131)	(0.392)	(0.151)	(0.367)
Observations	144	143	136	136	136	135

Note: Coefficients from OLS regressions. Dependent variable: Compete (0 Piece-Rate in Stage 3, 1 Tournament in Stage 3). Female (0 Male, 1, Female) indicates the agent's gender for *GenderInfo* and *NoGenderInfo* and the decision-maker's gender for *OwnDecision*. Perf. 2 - Perf. 1: Agent's/decision-maker's performance in Stage 2 minus the performance in Stage 1. Perf. 2: Agent's/decision-maker's performance in Stage 2. UK resident: the agent's/decision-maker's residency (0 Other, 1 UK resident). Age: the agent's/decision-maker's age in years. Risk: principal's/decision-maker's willingness to take risk measured with a lottery. The higher Risk, the more risk-loving is a person. Belief 2: principal's/decision-maker's belief about how Perf. 2 ranks within the group (1 = best to 4 = worst). Self-comp.: principals' stated choice for themselves in Stage 3 (0 Piece-Rate, 1 Tournament). Due to a technical problem, one subject in *OwnDecision* did not state their belief about their Stage 2 performance. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 3. Tournament entry by treatment

The baseline regressions by treatments (columns 1, 3 and 5) contained in Table 3 confirm the conclusions of the non-parametric tests: the negative and significant coefficient of *Female* (column 1) represents the GGC when subjects decide for themselves, whereas the GGC

is absent when principals decide (see columns 3 and 5).[15] We proceed by expanding the basic regressions in columns 2, 4, and 6. In *OwnDecision*, we add the decision-maker's characteristics and the belief about their Stage 2 rank (*Belief 2*). For *GenderInfo* and *NoGenderInfo*, we add the information the principal had about the agent when choosing on the agent's behalf and the beliefs about the agent's Stage 2 rank. Again, we observe a GGC in *OwnDecision* but not in the other two treatments. In line with much of the previous literature, *Female* (capturing the GGC) in *OwnDecision* decreases once beliefs and risk preferences are accounted for, and its statistical significance drops to the 7% level. Additionally, the data suggest that the beliefs about the relative Stage 2 performance are a strong predictor of higher tournament entry rates. We discuss the role of confidence in the agent's tournament performance further in Section 4.3. Our first main finding can be summarised as follows:

Result 1 *When deciding for themselves, fewer women than men choose to enter the tournament. However, we find no gender gap when principals make the tournament entry decisions on behalf of agents.*

Next, we examine the gender composition in the Stage 3 tournament and whether it differs between treatments. We find that the share of women competing in *GenderInfo* is greater than in *OwnDecision* (58.82% vs 34.72%; $\chi^2(1)$, $p < 0.01$), while the share of males competing is the same (55.88% vs 55.56%; $\chi^2(1)$, $p = 0.969$). To further explore these results, we run regressions that compare the outcomes across treatments, shown in Table A.3 (columns 1 and 2), including controls for performance, risk, beliefs and socio-demographics. The joint coefficient tests stated in row II of Table A.3 on the restriction GenderInfo + Female *x* GenderInfo = 0 ($p < 0.05$) reveal a significantly greater share of women competing. In contrast, the insignificant coefficients of *GenderInfo* show no difference for men competing in *GenderInfo* relative to *OwnDecision*. The interaction term *Female* x *GenderInfo* captures the difference in the GGC between treatments *OwnDecision* and *GenderInfo*. This term is sizeable, but at the threshold for statistical significance. In column 1, the gender gap appears

[15] To ease the interpretation of coefficients, we conduct Ordinary Least Squares regressions (OLS) throughout the analysis although the dependent variable is binary. Probit regressions lead to qualitatively the same results.

to be significantly different between the two treatments ($p = 0.053$), but this is no longer the case when we include the full set of controls in column 2 ($p = 0.173$).

Result 2 *More women compete when the decision to enter the tournament is made on their behalf. However, no difference is found for men.*

To analyse how information about the agent's gender affects tournament entry rates, we compare *GenderInfo* to *NoGenderInfo* and observe that the overall frequency of those competing in *NoGenderInfo* is lower compared to *GenderInfo* (45.59% vs 57.35%; $\chi^2(1)$, $p < 0.05$). This difference remains significant in the regressions in Table A.3 column 5 (see the coefficient of *NoGenderInfo*). This increase is mostly due to the share of males competing in *GenderInfo* being greater than in *NoGenderInfo* as indicated by a non-parametric test (55.88% vs 39.71%; $\chi^2(1)$, $p < 0.10$). For women we do not find support for an increase (58.82% vs 51.47%; $\chi^2(1)$, $p = 0.389$). The results from the non-parametric tests are in line with the coefficients of *NoGenderInfo* and the insignificant joint coefficient tests IV in Table A.2, columns 3 and 4. Nevertheless, inspecting Figure 1 suggests that this does not necessarily mean that principals in *GenderInfo* treat men and women differently. Rather, this may be due to the slightly, but insignificantly, varying tournament entry rates of women and men in *NoGenderInfo*, which is possibly the result of chance.

Result 3 *Informing the principal about the agent's gender leads to higher overall tournament entry rates.*

While we did not hypothesise Result 3 ex-ante, it may be worth taking a closer look at other reasons that may explain it. First, disregarding the principal's gender, we can rule out that the overall higher tournament entry rate in *GenderInfo* is caused by principals being on average more competitive than their counterparts in *NoGenderInfo*. When comparing the principal's competitiveness, measured by the incentive scheme they stated they would have chosen for themselves, we find no difference between treatments ($\chi^2(1)$, $p = 0.957$). Moreover, the tournament entry rate is higher in *GenderInfo*, even after we control for the principal's own competitiveness (see Table A.3, column 5). Second, one possible explanation may

be that information about the agent's gender changes the focus of the principals' decision. Comparing regressions 4 and 6 in Table 3 shows some support for this argument: all coefficients are comparable except for *Self-comp.* and *Risk*. It seems, then, that the principal's competitiveness (see *Self-comp.*) influences their decision when they know the agent's gender. However, *Risk* is a significant driver for the principal's decision only in *NoGenderInfo*.[16]

4.3 Determinants of the principal's choice

4.3.1 Principal's characteristics

To better understand the driving forces behind the principal's choices, we start by examining the role of age, UK residency and gender. For simplicity, we split the analyses in this section by treatments (see Table A.4 for *GenderInfo* and Table A.5 for *NoGenderInfo*). The first insights gained from the regression tables are that the GGC remains insignificant when controlling for a variety of principal characteristics (for the GGC in the respective treatments, see the coefficient of *(A) Female* in columns 1 and 2). For example, the principals' gender, age and UK residency do not affect their tournament decisions for the agents. We find no indication that either male or female principals choose the tournament compensation scheme at different rates (51.11% vs 51.82%; $\chi^2(1)$, $p = 0.906$, and insignificant coefficients *(P) Female*). These results persist when we distinguish between principals' and agents' gender (see Figure A.3): principals in *GenderInfo* do not condition their choice on the agent's gender (male principals: 57.14% vs 57.58%; $\chi^2(1)$, $p = 0.971$; female principals: 54.55% vs 60.00%; $\chi^2(1)$, $p = 0.649$, and insignificant coefficients *(A) Female* in Table A.4 columns 1 and 2).

We continue by testing whether the principal's risk preferences predict their choice of payment scheme. According to the existing literature, risk-averse subjects are less likely to compete, and females tend to be more risk-averse than men (see, e.g., Croson & Gneezy, 2009). Still, in terms of gender differences in choosing risk for others, Slovic et al. (1967) suggest no differences. We find only an effect of risk preferences in *NoGenderInfo*. Table

[16] Investigating the causes of Result 3 in more detail is beyond the focus of this paper, but we consider it an interesting question to be answered by future research.

A.5 shows that risk-seeking principals are more likely to choose the tournament for their agent (see *(P) Risk* in column 1). Our male principals are generally less risk-averse than their female counterparts in *NoGenderInfo* (Wilcoxon rank-sum test, $p < 0.05$) but not in *GenderInfo* (Wilcoxon rank-sum test; $p = 0.401$). These differences translate into a heterogeneous effect on making agents compete in *NoGenderInfo*. For male principals, see the coefficients of *(P) Risk*, and for female principals, see the joint coefficient tests I in column 2 of Tables A.4 and A.5 respectively.

Moving on, we turn to the effect of the principal's stated competitiveness (i.e., what they would have chosen for themselves) on their choice for the agent. Competitive principals are on average more likely to choose the tournament for their agent compared to non-competitive principals in *GenderInfo* (71.67 % vs 46.05%; $\chi^2(1)$, $p < 0.01$) but not in *NoGenderInfo* (53.33% vs 40.00%; $\chi^2(1)$, $p = 0.122$). These observations are corroborated by the regressions 1 contained in Tables A.4 and A.5 (see *(P) Self-comp.*).[17]

We also investigate whether gender differences in the principals' competitiveness affects their choice for the agent. Although a vast literature has found that women dislike performing in competitive environments, they do not necessarily mind choosing tournaments if they do not have to perform themselves, as shown by Niederle and Vesterlund (2007). Thus, it is not clear a priori in which direction the interaction between the principal's gender and competitiveness will pull our results. Recall that male principals are on average more likely than female principals to choose the tournament for themselves in *NoGenderInfo* (59.00% vs 30.43%; $\chi^2(1)$, $p < 0.01$), but not in *GenderInfo* (45.59% vs 42.65%; $\chi^2(1)$, $p = 0.730$). This raises further questions about whether and how these differences influence the tournament entry rates. Therefore, in columns 2 of Tables A.4 and A.5, we control for gender differences in the principal's competitiveness by including the interaction term *(P) Female x (P) Self-comp.*, observing no significant interaction effects.[18] Hence, the above

[17] We thank an anonymous referee for the suggestion to run, for each treatment, separate regressions for competitive and non-competitive principals. These regressions can be found in Tables A.3 and A.4, columns 3 and 4. We do not have strong evidence that the tournament entry rates differ significantly by the agent's gender in either treatments.

[18] We thank an anonymous referee for pointing out that it would also be interesting to report whether our results change when breaking down our previous analysis between competitive and non-competitive principals. Therefore, we also added columns 3 and 4 to Tables A.4 and A.5, which develop this aspect further. We find no

analysis shows that, even after controlling for principals' characteristics, their willingness to take risk or how competitive principals are, we still observe no GGC.

4.3.2 Principal's evaluation of the agent's relative performance

Another factor that may influence the principals' choice of payment scheme is their evaluation of the agents' performance. Previous research suggests that performance in a stereotypically male task tends to be evaluated differently between men and women (Egan, Matvos, & Seru, 2017; Sarsons, 2017; Reuben, Sapienza, & Zingales, 2014; Spencer, Steele, & Quinn, 1999). To reiterate, principals receive information about the total number of sums correctly solved by the agent in Stage 1 and Stage 2. To determine the importance of confidence in an agent's *relative* performance, we incentivise them to guess how well their agent's performance ranked compared to others in their session (Stage 1) and group (Stage 2) on a scale from top [1] to bottom [4] quartile.[19]

First, we analyse whether the male and female agent's relative tournament performance is evaluated differently. Hence, we focus on *GenderInfo* as this is the treatment in which principals are informed of the agent's gender. In Table 4, we report the principals' guessed rankings, broken down by the agent's and the principal's gender. At the aggregate level, we find no evidence that male and female agents' relative tournament performance is evaluated differently by the principals ($\chi^2(3)$, $p = 0.480$). In particular, principals guessed that 35% of male agents were the best in their group compared to 47% of female agents. Further, we analyse the interplay between the principal's and agent's gender. The evaluations of male and female agents do not differ for male ($\chi^2(3)$, $p = 0.104$) or female ($\chi^2(3)$, $p = 0.938$) principals. We employ the regression in column 5 of Table A.4, which includes an interaction between the agent's gender and the principal's belief about the Stage 2 performance. *Belief 2* remains a significant predictor for the principal's tournament decision for male and female agents (see the joint coefficient test III). The insignificant interaction term confirms that principals do not evaluate the performance differently depending on the agent's gender. Thus,

evidence of a GGC for competitive and not competitive principals (see coefficients of *(A) Female*).

[19] Confidence may be an important component of competitiveness, as suggested by van Veldhuizen (2021) and Gillen, Snowberg, and Yariv (2019).

principals do not have different beliefs about the rank of male and female agents. However, when evaluating their own Stage 2 tournament performance in *GenderInfo* and *OwnDecision*, 44% of males guessed that they were the best ones in the group compared to 24% of women ($\chi^2(3)$, $p < 0.01$), which is in line with the existing literature on the GGC.

	Pooled			***Female principal***			***Male principal***		
	Agent			**Agent**			**Agent**		
Rank	*Male*	*Female*	Total	*Male*	*Female*	Total	*Male*	*Female*	Total
1	24	32	56	15	14	29	9	18	27
2	30	24	54	13	14	27	17	10	27
3	10	10	20	4	6	10	6	4	10
4	4	2	6	1	1	2	3	1	4
Total	68	68	136	33	35	68	35	33	68

Note: The belief in Stage 2 is the principal's belief about how the agent's performance in Stage 2 ranks within the group (1 = best to 4 = worst).

TABLE 4. Principals' beliefs about agents' relative performance in Stage 2 in *GenderInfo*

Second, we assess how the principals' confidence in the agents' relative Stage 2 performances, independent of the agent's gender, affects tournament entry decisions. In *GenderInfo* and *NoGenderInfo*, the better an agent is perceived to have performed in the tournament compared to other group members ceteris paribus, the more likely he or she is to compete ($\chi^2(1)$ tests, $p < 0.01$). The regressions in Tables A.4 and A.5 support these findings (see coefficient *Belief 2*), confirming that the principals' beliefs about the agent's performance are a strong and significant predictor for the decision to make the agent compete. Columns 3 and 4 distinguish between competitive and not competitive principals, and show that the coefficients of *Belief 2* remain negative but it turns insignificant only for not competitive principals in *GenderInfo*. The main result relating to the closure of the GGC is robust to controlling for the principals' guessed ranking of the agent's relative performance. The regression results underline the robustness of no GGC in *GenderInfo* and *NoGenderInfo* when principals decide.

4.4 Efficiency

We proceed to examine whether tournament entry decisions are more efficient (i.e. better in terms of, e.g., expected payoffs, foregone earnings, or the performance of those in the tournament) when made by principals or by decision-makers. We divide the efficiency analysis into ex-ante measures (those based on Stage 2 performances) and ex-post measures (those based on the Stage 3 performances). For simplicity, we abstract all measures from gender differences and focus on treatment differences.[20]

4.4.1 Ex-ante efficiency

We start by assessing ex-ante efficiency, which is calculated based on Stage 2 performance.[21] First, inspired by the approach of Niederle and Vesterlund (2007), we compute the expected earnings under the piece-rate and tournament scheme using the subjects' Stage 2 performance to predict their future performance.[22] In our context, earnings are maximised if a subject competes when the expected earnings are higher in the tournament and stays out of the tournament when the expected earnings are higher under the piece-rate scheme. Table 5 reports the number of decision-makers and principals whose choices maximise expected earnings. The frequency of total maximised earnings differs significantly between treatments ($\chi^2(2)$, $p < 0.01$). Overall, decision-makers are better than principals at making earning-maximising choices. For completeness, Table 5 also includes the number of non-maximising choices, capturing those who either under- or over-compete. A person under-competes when it would be optimal to enter the tournament, but they do not compete; a person over-competes when it would be payoff-maximising to stay out of the tournament, but they still enter. The

[20] The results from the efficiency analysis split by treatment, role and gender are available on request. Please note that generally, gender does not play a role in the efficiency analysis.

[21] Stage 2 performance is employed as the predictor for future performances in Stage 3. We could have used Stage 3 performance to calculate expected earnings; however, some subjects compete in Stage 3, while others do not. Thus, Stage 3 performance is not strictly comparable, as it may vary with the underlying compensation scheme, as shown by Gneezy et al. (2003).

[22] The expected earnings under the piece-rate compensation scheme are calculated by multiplying the performance in Stage 2 by £0.5. We apply the following strategy when calculating the expected earnings under the tournament incentive: Within each group, we rank the subjects according to their Stage 2 performance from 1 (best) to 4 (worst). If a subject is ranked first or shares the first rank with another subject, the subject's performance in Stage 2 is multiplied with £2. If a subject is not ranked first, the expected tournament earnings are zero.

number of subjects who either under- or over-compete is lowest in *OwnDecision* and highest in *GenderInfo*.

	Maximising		**Not-maximising**		**Total**
	Tournament	***Piece-Rate***	***Tournament***	***Piece-Rate***	
OwnDecision	28	69	37	10	144
GenderInfo	24	44	54	14	136
NoGenderInfo	21	55	41	19	136
Total	73	168	132	43	416

TABLE 5. Number of decision-makers and principals who maximise and do not maximise expected earnings, by treatment

Second, we calculate the expected forgone earnings resulting from decision-makers and principals' payoff-inferior decisions across treatments. We compute the expected average forgone earnings as the difference between the expected earnings under the optimal decision and the earnings resulting from the principals or decision-maker's actual decision. The average forgone earnings are lowest in *OwnDecision* (2.67; $SD = 5.56$), followed by *NoGenderInfo* (4.33; $SD = 7.33$), and *GenderInfo* (4.35; $SD = 8.83$; Jonckheere–Terpstra test, $p < 0.05$). To interpret this result, recall that we generally observe higher tournament entry rates in the principal treatments. This may partially explain why we also find higher expected forgone earnings when principals decide compared to when subjects choose to compete themselves.

	Weak	***Intermediate***	***Strong***	**Total**
OwnDecision	15	28	22	65
GenderInfo	22	44	12	78
NoGenderInfo	15	36	11	62
Total	52	108	45	205

TABLE 6. Performance profiles of subjects competing in Stage 3, by treatment

Third, we investigate more closely how the decisions made by principals and decision-makers affect the performance profiles of competing subjects in Stage 3 (i.e. their expected performance quality). A similar approach was used by Brandts et al. (2015) and Balafoutas and Sutter (2012). We use aggregate performance quartiles from Stage 2 to classify subjects

into one of three performance profiles as follows: *weak* (less than 9 correct answers; first quartile), *intermediate* (9 to 13 correct answers; second and third quartiles) and *strong* (more than 13 correct answers; fourth quartile) performers. Table 6 indicates that there is a difference in the competitors' profiles between treatments ($\chi^2(2)$, $p < 0.10$). In particular, the number of strong-performing subjects is cut by nearly half when one is made to compete in *GenderInfo* and *NoGenderInfo* compared to when subjects choose to enter the Stage 3 tournament themselves (33.85% vs 15.38% vs 17.74%, $\chi^2(2)$, $p < 0.05$). Comparing performance profiles of agents competing in *GenderInfo* and *NoGenderInfo*, we find no significant difference ($\chi^2(2)$, $p = 0.842$).

To summarise, all three ex-ante measures point in the same direction: decision-makers make more efficient tournament entry decisions than principals.

4.4.2 Ex-post efficiency

In this section, we study efficiency from an ex-post perspective by evaluating whether being made to compete by someone else harms or improves the performance in Stage 3. In this respect, it is useful to distinguish between the performance of those competing and the performance of those winning, as suggested by Balafoutas and Sutter (2012).[23] Relatedly, Brandts et al. (2015) use the job market as an example to explain that one can evaluate efficiency concerns among the 'applicant pool', consisting of those who apply for a high-ranking job, and the 'winner pool', consisting of those who successfully obtain the job and win the tournament.

Specifically, we first compare the performances in *OwnDecision* to the treatments in which the principals choose. Table 8 shows that competing agents in *OwnDecision* solved on average more correct sums than agents being made to compete in *GenderInfo* and *NoGenderInfo* (Wilcoxon rank-sum tests, $p < 0.10$). Although these preliminary results suggest that the agents' performance is lower when being made to compete, we must consider the previous performance in Stage 1 and Stage 2 to measure ability. In the regressions presented in Table 7,

[23] Subjects win the tournament if their performance is higher than the Stage 2 performance of all other group members. This implies that for the ex-post efficiency analysis, similar to the ex-ante efficiency analysis, the mechanism of randomly choosing one winner if two subjects have the same performance does not apply.

we use Stage 3 performance as the dependent variable and control for past performance and decision-makers' or agents' characteristics (columns 1 and 2). Those who performed better in Stage 1 and Stage 2 also performed better when competing in Stage 3. We find that the negative effect of being made to compete is not robust to controlling for previous performance (see the coefficients of *GenderInfo* and *NoGenderInfo*).

	(1) Perf. 3	(2) Perf. 3	(3) Perf. 3	(4) Perf. 3
GenderInfo	-0.425	-0.444	-0.390	-0.423
	(0.411)	(0.433)	(0.522)	(0.530)
NoGenderInfo	0.095	0.069	0.339	0.146
	(0.442)	(0.477)	(0.605)	(0.657)
Perf. 2 - Perf. 1	-0.445 ***	-0.439 ***	-0.405 ***	-0.403 ***
	(0.083)	(0.085)	(0.112)	(0.117)
Perf. 2	1.094 ***	1.088 ***	1.074 ***	1.061 ***
	(0.043)	(0.045)	(0.053)	(0.064)
Female		-0.111		-0.443
		(0.337)		(0.473)
UK resident		-0.048		0.151
		(0.363)		(0.535)
Age		-0.023		-0.085
		(0.065)		(0.091)
Constant	0.223	0.889	1.213	3.403
	(0.551)	(1.698)	(0.822)	(2.548)
Observations	205	205	91	91

Note: Coefficients from OLS regressions. Dependent variable: Performance in Stage 3 of those competing (columns 1 and 2) and competing and winning (columns 3 and 4). The reference category in all regressions is *OwnDecision* for the dummy variables GenderInfo and NoGenderInfo, respectively. Perf. 2 - Perf. 1: Difference between the agent's performance in Stage 1 and Stage 2. Perf. 2: Agent's performance in Stage 2. Female: the agent's (*GenderInfo* and *NoGenderInfo*) and the decision-maker's gender (*OwnDecision*) (0 Male, 1 Female). UK resident: the agent's/decision-maker's residency (0 Other, 1 UK resident). Age: the agent's/decision-maker's age in years. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 7. Ex-post efficiency

Restricting the analysis to the winner's performance (see Table 8), our results suggest that winners in *OwnDecision* and *NoGenderInfo* perform significantly better than in *GenderInfo* (Wilcoxon rank-sum tests, $p < 0.05$). However, once we run the regressions reported in Table 7 and control for, among other things, past performance (see columns 3 and 4), we again observe that this effect is not robust (see coefficients of *GenderInfo* and *NoGenderInfo*).

	Competing	*Competing and winning*
OwnDecision	12.77 (4.76)	15.05 (3.86)
GenderInfo	11.13 (4.40)	13.96 (4.96)
NoGenderInfo	11.69 (5.30)	15.65 (4.65)

Note: The average number of correctly solved addition tasks in each treatment is shown with standard deviation in parentheses.

TABLE 8. Stage 3 performance of those competing and winning, by treatment

Finally, we focus on *GenderInfo* and *NoGenderInfo*. We ask whether Stage 3 performance is higher when the agent's stated choice coincides with the choice made by their principal. We pool *GenderInfo* and *NoGenderInfo* and observe that, out of all agents competing, only 40% stated they would have chosen to compete if given a choice. Those 40% of agents made to compete have on average a higher performance in the Stage 3 tournament than the remaining 60% of agents, whose choice would have been the piece-rate (12.77 vs 10.45; Wilcoxon rank-sum test, $p < 0.05$). Moreover, according to the non-parametric tests, they were more likely to win the tournament ($\chi^2(1)$, $p < 0.10$). We check the sensitivity of our results using regressions in Table A.6 with the Stage 3 tournament performance as the dependent variable. Whenever the principal's choice coincides with the agent's stated choice (see *Coincide*), we detect an effect significant at the 10% level, in column 2. Controlling for past performances, we again observe that better performers in Stage 1 and Stage 2 also perform better in the Stage 3 tournament. Overall, past performance matters more for Stage 3 performance than if the principal's choice and the agent's stated choice coincide. Taken together, our ex-post efficiency analysis does not suggest a negative effect of being made to compete on performance, in contrast to the ex-ante efficiency measures.

5 Conclusions

Over the last decades, much attention has been devoted to studying gender differences in competitiveness (for a summary, see Niederle, 2017). In response to these gender differences, there is a widespread sentiment from third parties that women should compete more and some individuals are even willing to pay for women to enter competitive environments more often

(Exley et al., 2020). Although most decisions in life are made by individuals themselves, many may be influenced by third parties, be they principals, managers, parents, superiors or societies. Despite this, surprisingly little is known about how individuals would make competitive decisions for others if given the opportunity. To investigate this uncharted area, this paper studies experimentally how competitive decisions are made on behalf of others.

Our paper offers three main results: First, whenever a principal decides for an agent, we find no gender difference in tournament entry as opposed to when individuals choose for themselves. In other words, our results indicate that, in situations when a principal decides for an agent, the gender difference in the willingness to compete does not translate to the the decisions of principals. Second, this result is mainly driven by more women being made to compete when someone else decides for them. Third, comparing only two environments in which principals decide, we observe more agents being made to compete when the agent's gender is known than when it is unknown.

In terms of determinants that may underlie the principals' decision, we document that risk-seeking and competitive principals are more likely to make their agent compete; however, this effect does not depend on the principal's or the agent's gender. Another factor that affects whether agents are made to compete are the principals' beliefs about their agent's relative tournament performance. Interestingly, while we document a gender gap in confidence among those performing in the competition, there is no difference in the principals' performance evaluation between female and male agents. We finalise our analysis by exploring ex-ante and ex-post efficiency to investigate whether principals or decision-makers make 'better' choices, as judged by different measures. Concerning ex-ante efficiency, principals' competition choices lead to fewer payoff-maximising outcomes, than when subjects decide for themselves. For ex-post efficiency, the overall tournament performance and winners' performances are lower when someone is made to compete; however, this effect is not robust to controlling for agents' previous performance.

Considering the accumulated evidence on the GGC, it is important to know whether these gender differences in tournament entry extend to environments in which the decision to compete is not made by individuals but by others who face the same monetary consequences

from tournament entries. Moreover, there is an increasing interest among researchers and policy-makers to understand the determinants of gender-specific preferences. We contribute to this conversation by investigating the causal role of gender information when choosing for others in the competitiveness domain. Our design allows us to observe how men and women make decisions for others, informing us about the generalisability of preferences for competition.

Future studies could fruitfully extend our experimental design to a variety of settings. For instance, in our experiment, principals do not know what their agent would prefer. An interesting question would be whether the GGC would reappear if principals are informed about the agents' preferences. In general, the preferences of agents can be very important in a contexts in which decisions are made on their behalf. It is likely that at least some individuals would experience a disutility when being forced to compete (or to take the piece-rate scheme) against their will. Our experiment is not equipped to quantify this kind of non-monetary disutility, but it may be important to consider in practice, along with potential ethical concerns of imposing a certain kind of choices on individuals.[24] In addition, considering a situation in which individuals decide if they want to delegate the tournament entry decision to another person may reveal potential gender differences in delegation decisions. Finally, investigating settings in which the principal's earnings are not perfectly aligned with the agent's could offer insights about further determinants and the role of incentives underlying the decision to make others compete.

[24]Such considerations echo discussions on the ethical dimensions of nudging and paternalism; see, for instance, Thaler and Sunstein (2003).

A - Supplementary Tables and Figures

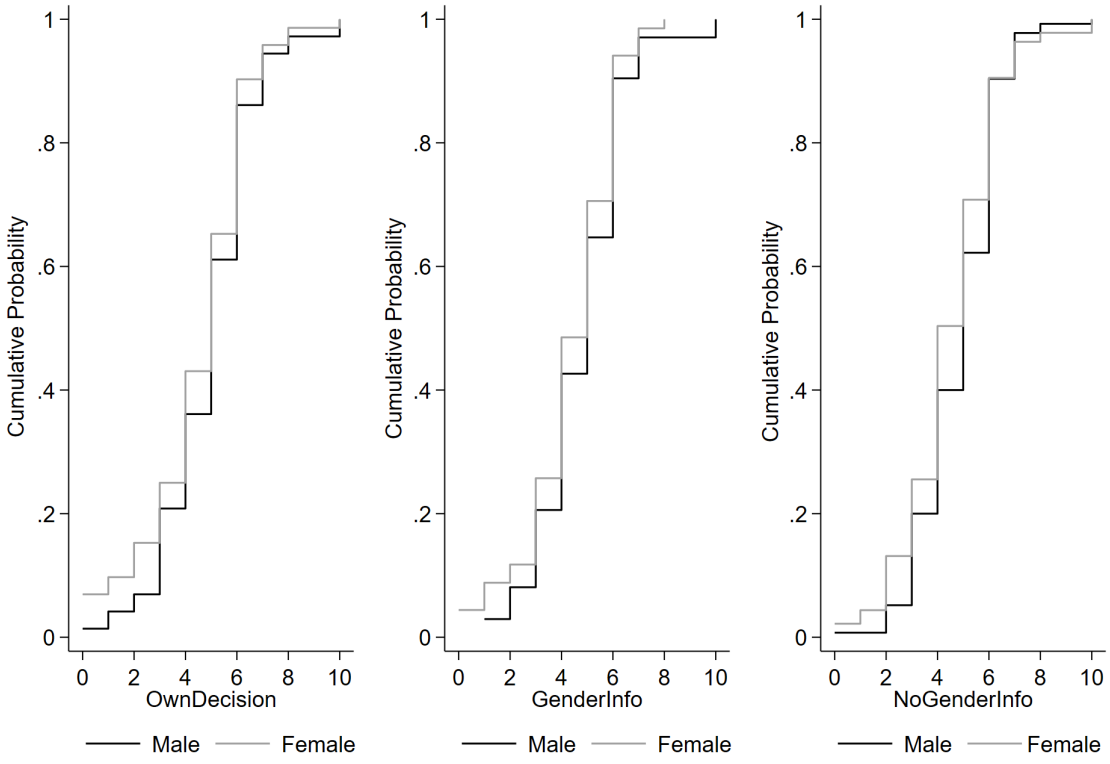
A.1 Descriptive statistics and performance in the task

	Pooled	*OwnDecision*	*GenderInfo*	*NoGenderInfo*	*p*-value
% Female	50.14	50.00	50.00	50.37	0.996
% UK resident	65.41	60.42	63.97	69.48	0.147
% Self-comp.	42.50	45.14	40.44	43.17	0.627
Age	21.83 (3.62)	22.38 (3.57)	21.97 (3.52)	21.39 (3.71)	0.001
Risk	4.65 (1.75)	4.73 (1.94)	4.58 (1.73)	4.67 (1.68)	0.617
Payoffs	10.04 (9.87)	10.66 (10.99)	10.30 (10.57)	9.45 (8.403)	0.994
N	688	144	272	272	

Note: Female: 0 Male, 1 Female. UK resident: 0 Other, 1 UK resident. Self-comp.: 0 Piece-rate, 1 Tournament. In *OwnDecision*, the measure captures the tournament entry choice of the decision-maker. In *GenderInfo* and *NoGenderInfo*, Self-comp. captures whether the agents and principals would have chosen to compete for themselves. Age is reported in years. Risk: willingness to take risk measured with a lottery. The higher Risk, the more risk-loving is a person. Payoffs are reported in £. The *p*-values are from Kruskal–Wallis and χ^2 tests, depending on the variable being continuous or not. The tests compare the values across the three treatments.

TABLE A.1. Descriptive statistics of subjects' characteristics by treatment

FIGURE A.1. CDFs of risk preferences in the lottery by gender and treatment



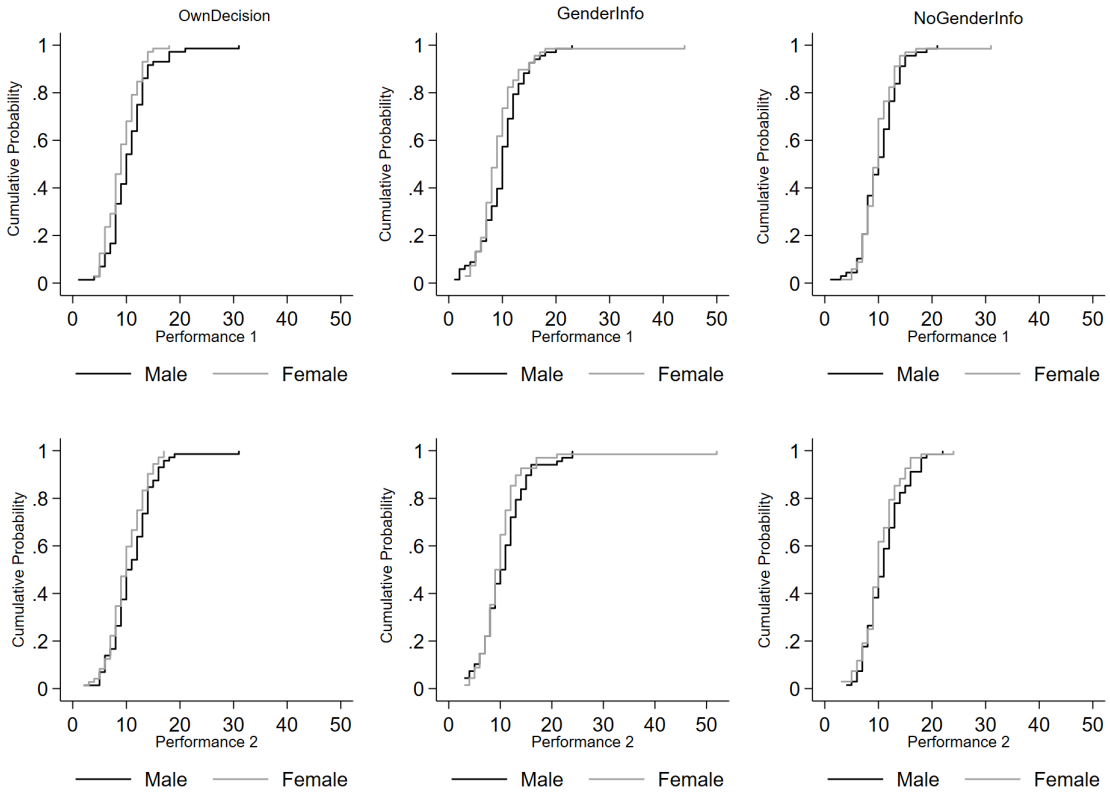
Note: The figures show the cumulative distribution functions for the risk preferences across all roles. The higher the risk variable, the more risk-seeking is a subject.

	GenderInfo		*NoGenderInfo*	
	Piece-rate	Tournament	Piece-rate	Tournament
Male (P)	29	39	37	30
Female (P)	29	39	37	32
Total	58	78	74	62
p-value	1.000		0.851	

Note: Reported p-values from χ^2 tests compare the number of principals by gender and payment-scheme for each treatment.

TABLE A.2. Number of principals choosing piece-rate and tournament for their agent, by the principal's gender and treatment

FIGURE A.2. CDFs of the Stage 1 and Stage 2 performances by gender and treatment



Note: The figures show the cumulative distribution functions for the number of correctly solved additions by agents in *GenderInfo* and *NoGenderInfo* and decision-makers in *OwnDecision*.

A.2 Gender gap in competitiveness

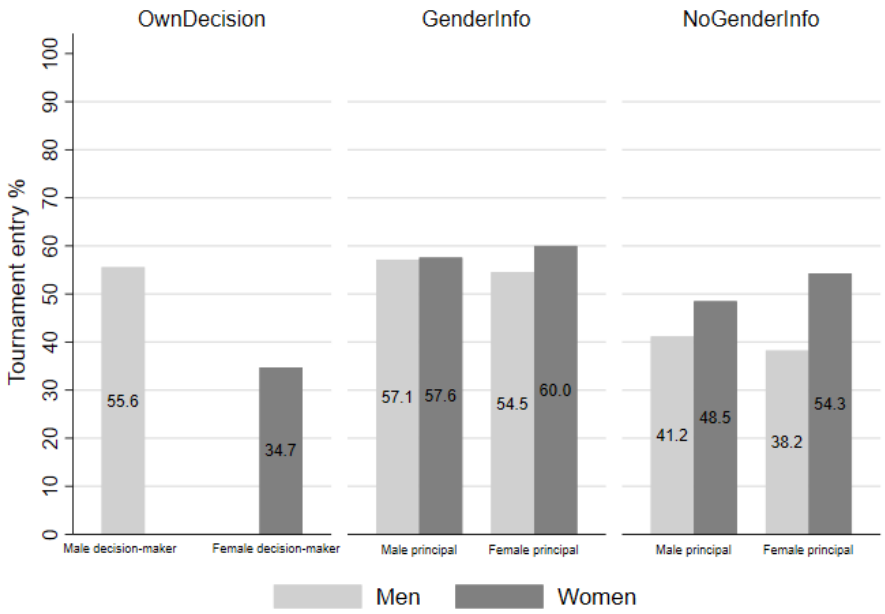
	OwnDecision vs. GenderInfo		*GenderInfo vs. NoGenderInfo*		
	(1)	(2)	(3)	(4)	(5)
	Compete	Compete	Compete	Compete	Compete
Female	-0.193**	-0.140*	0.029	0.034	
	(0.081)	(0.077)	(0.086)	(0.080)	
GenderInfo	0.009	0.008			
	(0.084)	(0.083)			
Female *x* GenderInfo	0.228*	0.158			
	(0.117)	(0.115)			
NoGenderInfo			-0.165*	-0.146*	-0.117**
			(0.085)	(0.082)	(0.056)
Female *x* NoGenderInfo			0.097	0.058	
			(0.121)	(0.114)	
Perf. 2 - Perf. 1	0.003	-0.015	0.012	0.004	0.003
	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
Perf. 2	0.015	0.009	0.002	0.000	-0.000
	(0.011)	(0.008)	(0.008)	(0.007)	(0.007)
UK resident		-0.144**		-0.047	-0.054
		(0.062)		(0.063)	(0.062)
Age		0.008		-0.008	-0.008
		(0.009)		(0.010)	(0.009)
Risk		0.028*		0.038**	0.038**
		(0.016)		(0.018)	(0.018)
Belief 2		-0.149***		-0.185***	-0.189***
		(0.033)		(0.040)	(0.040)
Self.-comp.				0.170***	0.169***
				(0.056)	(0.056)
Constant	0.387***	0.538*	0.533***	0.848***	0.886***
	(0.131)	(0.284)	(0.107)	(0.269)	(0.263)
Observations	280	279	272	271	271
I	0.688	0.834			
II	0.004	0.045			
III			0.144	0.261	
IV			0.432	0.266	

Note: Coefficients from OLS regressions. Dependent variable: Compete (0 Piece-Rate in Stage 3, 1 Tournament in Stage 3). Female (0 Male, 1, Female) indicates the agent's (*GenderInfo* and *NoGenderInfo*) and the decision-maker's gender (*OwnDecision*). Columns 1 and 2 compare *GenderInfo* with *OwnDecision*, the latter being the reference category for the dummy variable GenderInfo. Columns 3, 4 and 5 compare *NoGenderInfo* with *GenderInfo*, the latter being the reference category for the dummy variable NoGenderInfo. Perf. 2 - Perf. 1: Agent's/decision-maker's performance in Stage 2 minus the performance in Stage 1. Perf. 2: Agent's/decision-maker's performance in Stage 2. Risk: principal's/decision-maker's willingness to take risk measured with a lottery. Belief 2: the principal's/decision-maker's belief about how Perf. 2 ranks within the group (1 = best to 4 = worst). Self-comp.: principals' stated choice for themselves in Stage 3 (0 Piece-Rate, 1 Tournament). D Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values are shown for the following joint coefficient tests: I Female+Female *x* GenderInfo, II GenderInfo+Female *x* GenderInfo, III Female+Female *x* NoGenderInfo, and IV NoGenderInfo+Female *x* NoGenderInfo.

TABLE A.3. Tournament entry between treatments

A.3 Determinants

FIGURE A.3. Tournament entry by gender and treatment (N = 416)



Note: Stage 3 tournament entry of men and women separated by male/female decision-makers in (*OwnDecision*) and male/female principals in (*GenderInfo* and *NoGenderInfo*).

	(1) Compete	(2) Compete	(P) Not comp. (3) Compete	(P) Comp. (4) Compete	(5) Compete
(A) Female	0.056	0.058	0.025	0.166	0.071
	(0.086)	(0.087)	(0.138)	(0.121)	(0.204)
(A) Perf. 2 - Perf. 1	-0.005	-0.006	0.016	-0.030	-0.005
	(0.025)	(0.025)	(0.036)	(0.032)	(0.026)
(A) Perf. 2	0.006	0.005	-0.003	0.014	0.006
	(0.011)	(0.011)	(0.011)	(0.016)	(0.011)
(A) UK resident	-0.061	-0.060	-0.182	0.024	-0.060
	(0.090)	(0.090)	(0.145)	(0.113)	(0.090)
(A) Age	-0.010	-0.012	-0.012	-0.018	-0.010
	(0.014)	(0.015)	(0.022)	(0.022)	(0.014)
(P) Belief 2	-0.173***	-0.167***	-0.125	-0.242***	-0.169**
	(0.056)	(0.056)	(0.076)	(0.071)	(0.079)
(P) Risk	0.023	0.000	0.027	-0.005	0.022
	(0.027)	(0.041)	(0.038)	(0.043)	(0.027)
(P) Self.-comp.	0.225***	0.191			0.225***
	(0.085)	(0.122)			(0.084)
(P) Female	0.010	-0.223	-0.003	0.037	0.011
	(0.082)	(0.261)	(0.123)	(0.109)	(0.084)
(P) Age	-0.001	-0.001	0.001	-0.012	-0.001
	(0.010)	(0.010)	(0.013)	(0.019)	(0.010)
(P) UK resident	0.064	0.063	0.243	-0.159	0.065
	(0.098)	(0.099)	(0.171)	(0.105)	(0.099)
(P) Female *x* (P) Risk		0.043			
		(0.053)			
(P) Female *x* (P) Self.-comp.		0.078			
		(0.168)			
(P) Female *x* (P) Belief 2					-0.008
					(0.114)
Constant	0.831*	0.973**	0.775	1.685**	0.820*
	(0.442)	(0.468)	(0.577)	(0.671)	(0.450)
Observations	136	136	76	60	136
I		0.191			
II		0.024			
III					0.032

Note: Coefficients from OLS regressions. Column 3 reports the regressions for principals, who would chose the piece-rate themselves, i.e. Self-comp. = 0. Column 4 reports the regressions for principals, who would chose the tournament themselves, i.e. Self-comp. = 1. Dependent variable: Compete (0 Piece-Rate in Stage 3, 1 Tournament in Stage 3). Female (0 Male, 1, Female) indicates the agent's (A) or the principal's (P) gender. (A) Perf. 2 - Perf. 1: Difference between the agent's performance in Stage 1 and Stage 2. (A) Perf. 2: Agent's performance in Stage 2. UK resident: the agent's (A) or the principal's (P) residency (0 Other, 1 UK resident). Age: the agent's (A) or the principal's (P) age in years. (P) Risk: the principal's willingness to take risk measured with a lottery. The higher Risk, the more risk-loving is a person. (P) Belief 2: the principal's belief about how Perf. 2 ranks within the group (1 = best to 4 = worst). (P) Self-comp.: principals' stated choice for themselves in Stage 3 (0 Piece-Rate, 1 Tournament). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values are shown for the following joint coefficient tests: I (P) Risk+(P) Female *x* Risk, II (P) Self-comp.+(P) Female *x* (P) Self-comp., and III (P) Belief 2+(A) Female *x* (P) Belief 2.

TABLE A.4. The influence of principal characteristics and preferences on the principal's tournament choice in *GenderInfo*

	(1) Compete	(2) Compete	(P) Not comp. (3) Compete	(P) Comp. (4) Compete
(A) Female	0.069	0.056	0.033	0.111
	(0.088)	(0.088)	(0.131)	(0.141)
(A) Perf. 2 - Perf. 1	0.011	0.009	0.002	0.020
	(0.021)	(0.022)	(0.032)	(0.034)
(A) Perf. 2	-0.012	-0.014	-0.003	-0.022
	(0.011)	(0.012)	(0.016)	(0.017)
(A) UK resident	-0.055	-0.074	0.065	-0.262*
	(0.093)	(0.093)	(0.126)	(0.153)
(A) Age	-0.008	-0.010	-0.010	-0.003
	(0.013)	(0.013)	(0.012)	(0.025)
(P) Belief 2	-0.213***	-0.220***	-0.208**	-0.201**
	(0.059)	(0.058)	(0.085)	(0.097)
(P) Risk	0.054**	0.005	0.057*	0.043
	(0.023)	(0.036)	(0.034)	(0.034)
(P) Self.-comp.	0.112	0.151		
	(0.086)	(0.119)		
(P) Female	0.003	-0.341	0.008	-0.063
	(0.090)	(0.257)	(0.130)	(0.133)
(P) Age	-0.010	-0.010	-0.016	0.000
	(0.011)	(0.011)	(0.012)	(0.021)
(P) UK resident	-0.162	-0.170	-0.185	-0.191
	(0.106)	(0.108)	(0.174)	(0.169)
(P) Female *x* (P) Risk		0.080*		
		(0.046)		
(P) Female *x* (P) Self.-comp.		-0.084		
		(0.166)		
Constant	1.173**	1.497**	1.203*	1.247
	(0.556)	(0.579)	(0.719)	(0.901)
Observations	135	135	75	60
I		0.006		
II		0.577		

Note: Coefficients from OLS regressions. Column 3 reports the regressions for principals, who would chose the piece-rate themselves, i.e. Self-comp. = 0. Column 4 reports the regressions for principals, who would chose the tournament themselves, i.e. Self-comp. = 1. Dependent variable: Compete (0 Piece-Rate in Stage 3, 1 Tournament in Stage 3). Female (0 Male, 1, Female) indicates the agent's (A) or the principal's (P) gender. (A) Perf. 2 - Perf. 1: Difference between the agent's performance in Stage 1 and Stage 2. (A) Perf. 2: Agent's performance in Stage 2. UK resident: the agent's (A) or the principal's (P) residency (0 Other, 1 UK resident). Age: the agent's (A) or the principal's (P) age in years. (P) Belief 2: the principal's belief about how Perf. 2 ranks within the group (1 = best to 4 = worst). (P) Risk: the principal's willingness to take risk measured with a lottery. The higher Risk, the more risk-loving is a person. (P) Self-comp.: principals' stated choice for themselves in Stage 3 (0 Piece-Rate, 1 Tournament). Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. p-values are shown for the following joint coefficient tests: I (P) Risk+(P) Female *x* (P) Risk and II (P) Self-comp.+(P) Female *x* (P) Self-comp.

TABLE A.5. The influence of principal characteristics and preferences on the principal's tournament in *NoGenderInfo*

A.4 Robustness checks

	(1) Perf. 3	(2) Perf. 3	(3) Perf. 3	(4) Perf. 3
NoGenderInfo	0.522	0.522	0.659	0.439
	(0.369)	(0.371)	(0.481)	(0.492)
Perf. 2 - Perf. 1	-0.438***	-0.480***	-0.469***	-0.536***
	(0.096)	(0.099)	(0.116)	(0.119)
Perf. 2	1.181***	1.198***	1.163***	1.192***
	(0.053)	(0.054)	(0.057)	(0.071)
Coincide	-0.606	-0.725*	-0.268	-0.509
	(0.402)	(0.424)	(0.490)	(0.442)
Female		-0.228		-0.568
		(0.391)		(0.478)
UK resident		0.626		1.062*
		(0.412)		(0.545)
Age		0.032		0.023
		(0.096)		(0.126)
Constant	-0.891	-1.945	-0.119	-1.039
	(0.547)	(2.264)	(0.714)	(3.220)
Observations	140	140	60	60

Note: Coefficients from ordinary least squares regressions. Dependent variable: Performance in Stage 3 of those competing (columns 1 and 2) and competing and winning (columns 3 and 4). The reference category in all regression specifications is *GenderInfo* for the dummy variable NoGenderInfo. Perf. 2 - Perf. 1: Difference between the agent's performance in Stage 1 and Stage 2. Perf. 2: Agent's performance in Stage 2. Coincide: measures whether the agent's stated choice and the principal's actual choice coincide (1 coincide, 0 contradict). Female: the agent's gender (0 Male, 1 Female). UK resident: the agent's residency (0 Other, 1 UK resident). Age: the agent's age in years. Robust standard errors are reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

TABLE A.6. Stage 3 performance of competing and winning agents when the principal's choice coincides with the agent's stated choice

CHAPTER IV

Gender Effects in the Battle of the Sexes: a Tale of Two Countries*

*Co-authored with Fabrizio Adriani and Silvia Sonderegger.

Abstract

In a setting where inequality is ingrained and largely unavoidable (battle of the sexes), we investigate whether subjects condition their behavior on the gender of their opponent. In order to identify the role of culture and gender norms, we run the experiment in two countries, Norway and India, characterised by very different levels of gender inequality. Our evidence is consistent with different patterns of gender effects in the two countries. In India, subjects appear to become more 'hawkish' when facing a woman. This occurs only for lower education subjects and is observed primarily in female participants. Highly educated Indian participants do not discriminate between male and female opponents. In Norway, the data hint at a gender effect that is the opposite of what observed in India: it is present only in highly educated male participants, and takes the form of subjects appearing to become more 'hawkish' when facing a man. Our evidence suggests that these gender effects may be due to subjects experiencing different levels of inequality aversion depending on the gender of their opponent, in a manner that is mediated by their culture and gender norms.

JEL Codes: C70, C99, J16

Keywords: gender, social norms, coordination, culture, inequality aversion

1 Introduction

Many situations in life are characterised by intrinsic inequality. Examples include hierarchies in the workplace or leadership within a team, which is typically the domain of only one team member. How do people behave (and expect others to behave) when unequal outcomes are unavoidable? Do men and women adapt their behavior to the gender of their counterparts? What is the role of gender-related norms in guiding behavior? These are important questions that have so far been understudied.

This paper aims to answer these questions by examining behavior in a battle of the sexes game. A key feature of the battle of the sexes is that the players' choices either result in a payoff of zero for both (if the players fail to coordinate), or they generate positive payoffs (if the players coordinate) which, however, involve inequality: one player earns a higher payoff than the other. Clearly enough, each player would ideally like to earn the higher payoff, but this is not possible. If the players are to avoid earning zero, one must be willing to concede. Since the participants cannot communicate with each other, we expect that social norms and culture may be particularly relevant in guiding choices. To isolate the effect of gender, we employ a between subjects design where players are randomly allocated across control and treatment conditions. In the control condition, participants are matched with another player knowing only that he or she comes from their same country. In the treatment conditions, subjects know both the gender and the country of location of the other player. Behavior in this game depends on preferences (e.g., attitudes towards payoff inequality), as well as beliefs about the choices of others. We want to understand how gender enters into this picture and whether social norms play a role. In particular, the question we are interested in is whether our participants adapt their choice of action to the gender of their opponent.

Most of the literature to date has focused on identifying trait differences – e.g. in (over-)confidence, altruism, risk aversion, competitiveness etc. – between men and women in a fixed cultural context (see e.g. the surveys by Croson & Gneezy, 2009; Niederle, 2015). However, culture and underlying social norms or conventions are clearly very important in shaping behavior (Bicchieri, 2005, 2016; Charles, Guryan, & Pan, 2018; Jayachandran, 2015; Gelfand, 2019), and this paper aims to shed further light on this less-studied dimension of gender effects.[1] To characterise the role of culture, we run our online experiment in two

[1]The role of culture in shaping norms and behavior is also the focus of recent works looking at the role

countries, India and Norway, that feature very different gender norms and levels of gender inequality.[2]

There are several reasons why gender norms may matter in the game we consider. First, gender norms may act as a coordination device which focalises expectations about the equilibrium to be played. For instance, in societies where women have lower status and entitlement than men, subjects may naturally expect that everybody will coordinate on the man's preferred equilibrium (Cooper & Weber, 2020).

A second channel through which gender norms may matter is through their relationship with preferences. For instance, subjects may obtain different (dis-)utility from earning the lower payoff (500) when the opponent earning the higher payoff (1000) is a man compared to when it is a woman. These preferences may themselves be the result of an internalized social norm – see e.g. (Bicchieri, 2010) and (Postlewaite, 2011) for further discussion.[3] Intuitively, if in a society it is customary that men should obtain the upper hand whenever their interests clash with those of women, this might shape what people find acceptable or unacceptable. As a result, in a gender-unequal society, subjects may find it more acceptable to have their opponent come on top when their opponent is a man rather than a woman – a form of gender-dependent aversion to inequality. It is also possible that subjects of both genders may come to believe that it is morally 'right' that men should earn a higher payoff than women.[4]

Gender effects in the form of subjects conditioning their behavior on their opponent's gender have been documented by previous literature in contexts such as bargaining (Ayres & Siegelman, 1995; Eckel & Grossman, 2001; Sutter, Bosman, Kocher, & van Winden, 2009), dictator games (Ben-Ner, Kong, & Putterman, 2004), and ultimatum games (Solnick, 2001), always within a homogeneous cultural context.[5] Görges (2020) considers a lab experiment

of culture in shaping fairness views (see e.g. Cappelen, Falch, & Tungodden, 2020; Almås, Cappelen, & Tungodden, 2020).

[2]For instance, Norway is classed 2nd and India is classed 112th out of 153 countries in the most recent Global Gender Gap Index rankings. Source: World Economic Forum (2020)

[3]Classical choice theory treats preferences as an exogenously fixed black-box, but there is an increasing consensus that they may actually be more fluid and malleable than that. Bowles (1998), for instance, compares the process of preference formation to the acquisition of an accent. Recent empirical research argues that preferences respond to experiences such as cultural exposure (e.g., Atkin, 2013) and family influences (e.g., Fernández, Fogli, & Olivetti, 2004).

[4]A prominent theory in social psychology – justification theory – argues that, often, disadvantaged groups come to rationalise their lower status on objective grounds, presumably as part of a process of cognitive dissonance reduction (see e.g. the survey by Jost & Banaji, 1994).

[5]A related literature looks at whether individuals condition their behavior on their opponent's ethnic, sexual

where German (heterosexual) couples play a battle of the sexes in which the salience of gender roles is varied exogenously. The author's finding that women (but not men) respond strongly to gender roles salience is consistent with the observation by Croson and Gneezy (2009) that, across a wide range of experiments, women appear more sensitive to social cues when choosing their actions compared to men.

The work that is most related to ours is a study by Holm (2000). In a battle of the sexes game in which student participants from Sweden and the US are aware of the gender of their opponent, he finds that participants of both genders are more likely to behave hawkishly against women than against men.[6] The question is how to interpret this finding. As pointed out by the author himself, in this game, players are strongly incentivised to identify a focal point on which to coordinate. At the same time, they are also operating in a highly stylised environment where gender labels are essentially the only thing they can use as coordination device. This raises the possibility that gender effects may be driven primarily by coordination motives, and, as such, may bear little relation to underlying gender norms or gender inequality outside of the lab.

To address this and other issues, our design departs from Holm (2000) in a number of important respects. First, we elicit the subjects' expectations about the likely action of their opponent. This is done in order to distinguish between effects that are purely driven by coordination concerns as opposed to other mechanisms, such as aversion to disadvantageous inequality. Second, we include a treatment in which subjects are not informed of the gender of their opponent. This allows to identify any systematic differences (in preferences or beliefs) between men and women. Finally, our experiment includes a varied subject sample. We consider subjects from two countries (Norway and India) that exhibit very different levels of gender inequality.[7] Our subjects also differ in terms of other characteristics, such as education level, allowing us to paint a more nuanced picture of gender effects. Previous

or political affiliation (see, e.g. Fershtman & Gneezy, 2001a; Aksoy, Chadd, & Koh, 2021; Dimant, 2021).

[6]This result is, however, not replicated by Boschini et al. (2014), who find no gender effect in their study involving a large sample of the Swedish population.

[7]We argue that the difference in gender inequality between India and Norway nowadays is greater than the difference between Sweden and US (the two countries analysed in Holm's study) around the year 2000. For instance, the world value survey reports that the share of respondents disagreeing with the statement "When jobs are scarce men should have more right to a job than women" in 1999 was 93% for Sweden and 81.3% for the US. These numbers are very comparable and are actually similar to the numbers we have for 2017-20 Norway, where 91.3% of respondents disagreed. In contrast, the most recent wave available for India (2012) reports that only 22.6% of respondent disagreed with the statement, while 55.4% agreed with it (17.1% were unsure).

studies have shown that people who are more educated tend to hold more egalitarian views on gender– see e.g., the influential early studies by Thornton, Alwin, and Camburn (1983) and Kane (1995), as well as more recent evidence by Du, Xiao, and Zhao (2021). Outside of economics, the relationship between education and gender attitudes has been the object of intense debate – as attested for instance by the large literature in other social sciences, such as sociology.[8] In experiments like ours – where participants are heterogeneous in their education levels – it is therefore essential to look at how subgroups with different educational backgrounds behave when facing a man or a woman.

Our data present a number of interesting patterns. First, in the more gender-unequal society (India), subjects without university education are more likely to behave hawkishly (i.e., choose the action corresponding to their preferred equilibrium) when facing a woman. In Norway, on the other hand, this type of gender effect does not arise, if anything, we observe some evidence of the opposite, as discussed below. Second, the gender effect appears to be strongest among *Indian women* with lower education. In this subgroup, the probability of selecting the hawkish action when matched with another woman is 20% higher than when matched with a man. Indian men without university education are also more aggressive when facing a woman, but the effect is less pronounced. University educated Indians, on the other hand, exhibit no gender effect. These subjects behave in the same way independently of the gender of their opponent.

In Norway, the data suggest the *opposite* of what observed in India: *university educated Norwegian men* appear more likely to behave hawkishly when facing a *man*. Interestingly, gender effects in both India and Norway do not seem to be driven by the expectation that male and female opponents are likely to choose different actions. This suggests that our subjects do not simply use gender labels as a coordination device, as in that case the differential treatment of male and female opponents would be driven by expectations. Instead, the evidence is consistent with a preference-based mechanism. For instance, lower education Indian women may be more dovish when playing against a male opponent *not* because they expect male opponents to be more aggressive than females, but because they find a situation where the male opponent earns the highest payoff relatively more acceptable.[9]

[8] See e.g. Bolzendahl and Myers (2004); Brewster and Padavic (2000); Bryant (2003); Ciabattari (2001); Corrigall and Konrad (2007); Brooks and Bolzendahl (2004); Moore and Vanneman (2003); Tallichet and Willits (1986); Fan and Marini (2000); Cunningham, Beutel, Barber, and Thornton (2005).

[9] This shares similarities with the literature on identity. One of the most renowned papers in this literature is

Teasing out the different channels through which gender effects may operate can have important implications for policy interventions aimed at improving the condition of women in gender-unequal societies. Consider for instance a situation where women have internalised the dominant gender norm, coming to perceive their lower status as something 'natural', rather than as a state of affairs they would like to move away from. This could imply that policies that take the form of broadening up the range of options available to women may not be very effective unless they take place in conjunction with other interventions aimed at encouraging a stronger sense of female entitlement. Our design allows to shed light on the different drivers of gender effects.

The remainder of the paper is organised as follows. Section 2 discusses the experimental design and procedures, while Section 3 explains the theoretical rationales for our treatments and provides a series of conjectures. Section 4 is devoted to our empirical results, and Section 5 concludes the paper.

2 Experimental Design and Procedures

Our online experiment considers a canonical game that is particularly well-suited to study coordination in the presence of a conflict of interests: the battle of the sexes. Two players must simultaneously choose one of two actions, A or B. If both players select A, Player 1 earns 1000 points, and Player 2 earns 500, while if they both select B, Player 1 earns 500, and Player 2 earns 1000 points.[10] In what follows, we borrow Holm (2000)'s terminology and say that a player i 'plays hawkish' (or 'selects the hawkish action') if he or she chooses the action which, if both players were to coordinate upon it, would give i 1000 points, while player i 'plays dovish' (or 'selects the dovish action') if he or she chooses the action which would give i 500 points. If one player chooses A and the other chooses B, they both earn 0 points. The structure of the game is summarised below.

Akerlof and Kranton (2000), which formalizes the idea that a person's sense of self is tied to different social identities which affect individual behavior. This notion is also found in other disciplines such as sociology (Stryker, 1968; Burke & Tully, 1977; Burke, 1991) and anthropology (Erikson, 1994) but has been widely accepted in economics only in recent years. Clearly enough, different gender norms play an important role in constructing individual identity (Kranton, 2016).

[10] 1000 points = 60 NOK in Norway or 220 INR in India.

		Player 1	
		A	B
Player 2	A	$(1000, 500)$	$(0, 0)$
	B	$(0, 0)$	$(500, 1000)$

We employed a between subjects design. Each subject played the game only once. Participants were always paired with a participant from the same country and were randomly assigned to one of a series of treatments.[11] In the Gender Unknown treatment, players were informed that their opponent was from the same country but were not told about their gender (they only knew that he or she was male or female with equal probability). In the Gender Known treatments, players were told both their opponent's gender and that he or she was from their same country. This generated a total of three cases, depending on the gender of both players: (i) female-female matches, (ii) male-male matches and (iii) mixed matches (female-male).

In all treatments, participants were made aware that their opponent had received the same information as they did (namely, the country location and, in the Gender Known treatments, also the gender of the other player). To ensure that subjects paid sufficient attention to the gender of their opponent, the information about gender and country of location was displayed with a few seconds' delay, and subjects were asked to wait until the information appeared.[12] In the Gender Known treatment, subjects were informed that their opponent had an equal chance of being a man or a woman. Information about the location (India or Norway) was included to ensure that subjects were aware that their opponent came from the same culture – and, hence, had been exposed to similar gender norms. Including location information in the opponent's description also made the true focus of the study less clear.[13]

Table A.1 in the Appendix reports the number of participants across countries and treat-

[11] For practical reasons, participants were paired ex-post rather than in real time.

[12] Currently, there is no commonly accepted way to convey gender in online experiments. Possible options include informing subjects of their opponent's first name (Chakraborty, 2019), using avatars (Charness, Cobo-Reyes, Sanches, & Meraglia, 2020), or using pictures (Castillo & Petrie, 2010). We did not choose any of these options in order to avoid the possibility that first names, avatars, or pictures might provide social cues (e.g., about social class, religion, or caste), which could act as confounding effects.

[13] This raises the question of the possible roles of experimenter demand effects and social desirability bias in our data. We are not worried about social desirability as this is part of what we are trying to infer: differential attitudes towards men and women in India and Norway. In terms of experimenter demand effects, we should note that our subjects played only once (either against a woman or a man). As a result, we really did not give them any opportunity for gender discrimination in the commonly used sense (i.e. treat different genders differently). Any gender related difference in behaviour are obtained between subjects and there is very little subjects can do individually to align with the objectives of the experiment.

ments. A total of 1676 respondents took part in our online study 802 in India and 874 in Norway. The experiment took place between the months of November 2020 and January 2021 and was programmed with the Qualtrics online platform.[14] Subjects were recruited in both countries through the market research company Dynata, which is regularly used to conduct academic studies (e.g., see Haaland, Roth, & Wohlfart, 2020). The instructions were translated into both Hindi and Norwegian.[15] The experiment started with a number of questions on demographics, namely gender, age, education level (primary, secondary, or university), place of residence (village, small city, or big city), and religiousness. The information on gender was later used to assign subjects to the treatments. We also used the information on education to ensure sufficient within-treatment variation in the subjects' education levels. This was done in order to be able to isolate gender effects within different sub-populations (low/high education) in each country,

In the following stage, participants were informed about the main task and had to answer a series of control questions to check whether they understood the incentive structure of the study. Subjects were only given three attempts at answering these questions - those who failed to provide the correct answer for more than three times were excluded from the experiment. Once players passed the control questions, and after having been informed about their opponent's characteristics (either gender + country location or only country location, depending on the treatment) they were asked to state which action they expected their opponent to choose.[16] If they guessed correctly, the subjects received a small fixed payment on top of the final payoff. Incentivizing beliefs has been shown to increase accuracy – see, e.g. Gächter and Renner (2010) – but also raises the possibility that subjects might misrepresent their true expectations in order to hedge their bets. Blanco, Engelmann, Koch, and Normann (2010) look specifically into this possibility within the context of – among others – a battle of the sexes game. They compare a "hedging-prone" design where both the main task and beliefs are incentivised with a "hedging-proof" design where only one task is incentivised. Their findings suggest that, although possible in principle, hedging incentive

[14]The study was pre-registered in the AEA RCT Registry: https://www.socialscienceregistry.org/trials/6272

[15]For consistency, in India our data collection mainly focused on Hindi-speaking states (see Appendix B for a complete list). Since India has 22 official languages, translating the experiment in each of them would not have been feasible. We therefore focused on the most commonly spoken language, Hindi, but also gave the option to complete the experiment in English. Detailed instructions can be found in Appendix B.

[16]Our measure of expectations is thus relatively coarse. On the one hand, this implies that it could fail to fully account for heterogeneity in expectations. On the other hand, it has the advantage of being simple and easy to understand by all participants.

matter little in practice, especially when, as in our case, beliefs are elicited by simply asking for a subject's guess of the choice of the other player.[17] Another important observation is that, even if hedging bets were present and substantial in our experiment, this should affect all our treatments equally and would therefore not invalidate our results, which focus on how the opponent's gender affects behavior. We further discuss hedging in our Results section.

Once expectations were elicited, participants were presented with the main task. After they had chosen their action, they were then asked a series of follow-up questions, including why they chose one option over the other, whether they recalled the gender of their opponent – in order to verify that in the Gender Known treatments subjects had indeed paid attention to their opponent's gender – and what they thought was the purpose of the study.[18] Finally, we asked subjects to express their degree of agreement with a series of statements on gender roles and the position of women in the family, based on the World Value Survey (Inglehart et al., 2014). Average completion time for the whole study was 611 seconds (approx. 10 minutes) in India and 492 seconds (approx. 8 minutes) in Norway.[19]

2.1 Sample description

The descriptive statistics of our sample are reported in Table 1. As shown in the table, within each country, the profiles of men and women are similar.

In terms of cross-country comparisons, Norwegian subjects are slightly older (36 years on average, versus 30 years in India) and more likely to live in a village compared to their Indian counterparts (30% versus 6% in India). On the other hand, Indian subjects are, on average, more religious. In terms of education, the two samples are comparable: the share of university educated subjects is 60% in India and 50% in Norway.

There are however differences in terms of the representativeness of the sample. In India, our subjects are much more educated compared to the population at large: 60% are university educated, against 12% in the overall Indian population.[20]

[17] See also Nyarko and Schotter (2002) and Costa-Gomes and Weizsäcker (2008) for indirect evidence ruling out that hedging may play an important role when beliefs are elicited.

[18] The results are presented in Table A.2; 92% of Indians and 96% of Norwegians correctly recalled their opponent's gender. Concerning the objective of the study, less than 10% of male or female subjects in either country stated that they thought gender was the main focus of our study. Finally, in the Gender Unknown treatment, we asked subjects to guess the gender of their opponent. In both countries, the share of subjects who believed their opponent was male was about 50%.

[19] Median completion times for India was 509 seconds while for Norway it was 379 seconds.

[20] Furthermore, in our sample, the share of Indian participants who completed only primary-level education

	(1) Overall Mean (St.Dev.)	(2) India (F) Mean (St.Dev.)	(3) India (M) Mean (St.Dev.)	(4) Norway (F) Mean (St.Dev.)	(5) Norway (M) Mean (St.Dev.)
Demographics					
Age*	33.14 (12.77)	31.31 (10.53)	27.94 (9.80)	35.14 (13.19)	38.92 (14.59)
Country national	94% (.)	100% (.)	99% (.)	87% (.)	91% (.)
Education					
Primary	4% (.)	1% (.)	2% (.)	7% (.)	5% (.)
Secondary	43% (.)	37% (.)	40% (.)	44% (.)	51% (.)
University degree	54% (.)	62% (.)	58% (.)	49% (.)	45% (.)
Religiousness					
Not Religious	28% (.)	7% (.)	10% (.)	49% (.)	51% (.)
Moderately religious	53% (.)	58% (.)	63% (.)	46% (.)	45% (.)
Very religious	18% (.)	35% (.)	27% (.)	4% (.)	5% (.)
Live					
Village	18% (.)	4% (.)	9% (.)	29% (.)	31% (.)
Small city	29% (.)	23% (.)	25% (.)	34% (.)	35% (.)
Big city	53% (.)	72% (.)	66% (.)	37% (.)	34% (.)
Main variables					
Expect Hawkish	70% (.)	69% (.)	70% (.)	74% (.)	66% (.)
Chose Hawkish	50% (.)	52% (.)	52% (.)	48% (.)	49% (.)
N	1678	427	449	403	399

Notes: Summary statistics of the participants. Variables are continuous when asterisked and binary otherwise. Country national is a binary variable which takes value one if the respondent claimed to be a citizen of the country where the experiment took place and zero otherwise.

TABLE 1. Summary statistics

Hence, highly educated subjects are over-represented in our sample. Since education is known to affect gender attitudes, in the analysis that follows it is therefore important to fully account for it.[21]

Compared to the Indian subjects, our Norwegian sample is more representative. The share of university educated subjects is about 50% against a share of 44% in the whole population.[22]

3 Hypotheses

The game we consider entails a conflict of interests: although each player would ideally like to earn the 1000 points, this outcome is not possible for both of them. Someone has to concede if the players are to avoid earning zero points. The question is, who will choose

is very small, 1.5%, implying that our 'lower-education' subjects are relatively highly educated: almost all of them have secondary-level education. Indian subjects are also much more likely to live in a big city relative to the population at large – 69% of our Indian subjects, while across India, the share of urban population is only 35%. Sources: *World Data Atlas* (2020); Office of the Registrar General & Census Commissioner (2011).

[21] As we will see, our analysis shows that university educated Indian subjects adopt the same behavior when facing male and female opponents, while this is not the case for subjects with lower education level. Failing to account for education would therefore induce us to *underestimate* gender effects compared to what we would obtain in a more representative sample, with a larger share of low-educated individuals.

[22] Similarly, the urban population is 83% in the whole country and 70% in our sample. Sources: OECD. (2019); *World Data Atlas* (2020).

to concede and who will choose to stand their ground? Since players are not allowed to communicate, each needs to form a conjecture about what the other player will do. However, this depends on what the other player expects *them* to do. Consider, for instance, a mixed match in which the subjects are informed of each other's genders. If the man thinks that his female opponent expects him to choose the hawkish action, he is indeed inclined to play hawkish, as he anticipates that she is likely to play dovish. The opposite is also true. A man who thinks that his female opponent expects him to choose the dovish action finds it optimal to play dovish, since he anticipates that she will play hawkish.

The above discussion highlights a key feature of the game: the presence of multiple equilibria.[23] For a player, different actions are optimal depending on which equilibrium is being played. In the absence of explicit communication, subjects will use salient social cues as focal points to solve this coordination problem. While same-gender matches lack a natural focal point, in mixed matches, gender is a clear candidate to play this role (Holm, 2000): if subjects expect their opponent to follow the 'man = dominant, woman = submissive' stereotype, they will find it optimal to conform to it themselves. In this case, their behavior will differ depending on whether they are in a same-gender or a mixed-gender match. Following Cooper and Weber (2020), we conjecture that the salience of gender may increase according to the strength of gender norms in a society.

Another reason why we may observe gender effects is rooted in preferences. Preferences may interact with gender in two ways. First, the utility of a player from a given outcome may depend on the gender of the opponent. For instance, a player may be more or less averse to having their opponent come on top depending on the opponent's gender – a form of gender-dependent inequality aversion. This could for instance imply that earning the lower payoff of 500 may be more acceptable if the opponent earning 1000 is, say, a man rather than a woman. If this is the case then, keeping everything else equal, subjects will be more likely to select the hawkish action when facing a woman.[24] There is some experimental

[23] In the canonical case in which players are exclusively concerned with maximising own material payoff, the game has two pure strategy equilibria: (A, A) and (B, B). In each of these two equilibria, one player plays his or her hawkish action, and the other plays his or her dovish action. In addition, there is a mixed strategy equilibrium in which each player plays his or her hawkish action with probability $\frac{2}{3}$.

[24] In our game, inequality aversion introduces the possibility of new equilibria. In particular, the equilibrium (A, B) – where both players select their hawkish action and obtain a payoff of zero – will arise whenever the subjects are sufficiently averse to earning a lower payoff than their opponent. It is worth mentioning that, although in the text we primarily focus on disadvantageous inequality aversion, people may of course also be averse to *advantageous* inequality. However, in our data, aversion to advantageous inequality does not seem to matter much. This is further discussed in Section 4.

evidence backing the notion that social preferences may depend on the gender of the subject's opponent. In the context of a dictator game, Ben-Ner et al. (2004), find that female dictators give systematically less to other women compared to men, thus expressing weaker other-regarding preferences when facing a woman. In a similar vein, Solnick (2001) finds that, in the ultimatum game, receivers (of both genders) choose a lower minimum acceptable offer when facing a male sender. This suggests that, when the sender is a man, receivers become less averse to disadvantageous inequality.

Again, we expect this kind of gender-dependence to be affected by the subjects' cultural environment. Intuitively, in more gender-unequal societies, women are used to finding themselves in a disadvantaged position compared to men, and men are used to being advantaged compared to women. In mixed matches, we may therefore envisage that men and women will evaluate their utility relative to this inegalitarian reference point. Compared to same-gender matches, where – in the absence of other social cues – the natural reference point is egalitarian, this would result in women being less and men being more averse to disadvantageous inequality when facing a member of the opposite sex. Another (related) mechanism is that, in more gender-unequal societies, people may 'internalize' the dominant social norm and come to believe that the differential status of men and women is legitimate and justified on objective grounds, which could result in the pattern of inequality aversion described above (this is the idea behind justification theory in social psychology, discussed e.g. in Jost, 2019).

It is also possible that the utility from a given outcome may depend on the subject's *own* gender. The existing literature has identified a number of dimensions in which the preferences of men and women differ, most notably competitiveness and confidence – men are more competitive and more (over-)confident than women (e.g. see Niederle & Vesterlund, 2007; Niederle, 2015). Suppose, for instance, that men are inherently more 'aggressive' or more inequality-averse than women. This may help rationalize gender effects in the Gender Known treatment, since players facing a male opponent may be more reluctant to play hawkish compared to players facing a female opponent.

Our experimental design adds to previous literature by allowing to shed light on the drivers of gender effects.

· By eliciting the subjects' beliefs about their opponent's choice of action, we are able

to identify whether a gender effect is driven by subjects expecting men and women opponents to behave differently. As discussed, this is what we should observe when gender acts as a coordination device. For instance, if player 1 (male) and player 2 (female) use their gender to coordinate on the (1000, 500) equilibrium, then player 2 should expect player 1 to play hawkish and 1 should expect 2 to play dovish. If this is the underlying story, then any gender effect should disappear once expectations are accounted for: conditional on expecting the opponent to play hawkish or dovish, a player should select the same strategy regardless of the opponent's gender. If, on the other hand, the gender effect arises *even* when controlling for expectations, this points to the presence of more subtle effects. For example, if we observe that players who expect their opponent to behave hawkishly are more likely to choose the hawkish action when facing a woman (or a man), this suggests that the subjects' aversion to disadvantageous inequality may depend on the opponent's gender.

- The Gender Unknown treatment allows to investigate whether there are underlying differences in preferences between the genders, which may in turn generate gender effects. Suppose that we observe a gender effect in the form of subjects becoming more hawkish when facing a woman compared to a man. A possible reason for this is that women are inherently less aggressive than men and, hence, it is optimal to be more aggressive when facing a woman. The Gender Unknown treatment allows us to investigate this hypothesis by comparing the behavior of men and women when they are facing the exact same conditions. For instance, if we find that, in the Gender Unknown treatment (where they face the exact same conditions) men and women adopt the same behavior, this suggests that there are no inherent differences (e.g., in innate "hawkishness" or aggressiveness) between the genders, and, hence, any observed gender effect in the Gender Known treatment cannot be rationalized on these grounds. A recent survey by Falk and Hermle (2018) finds that the preferences of men and women differ more in more gender-equal societies. An interesting question is whether this prediction extends to our data.[25]

[25]The Gender Unknown treatment also allows to identify possible gender-specific biases in expectations about the opponent's likely action – e.g., one gender being systematically more 'optimistic' than the other. Existing literature has shown that men tend to be significantly more optimistic than women on issues such as economic outlook and financial markets (see e.g. Jacobsen, Lee, Marquering, & Zhang, 2014). In our context, optimism (or pessimism) might affect the beliefs about the likely action of a subject's opponent.

4 Results

We start with summary statistics. The last two rows of Table 1 report the share of hawkish choices depending on gender and country, showing that the aggregate propensity to behave hawkishly is comparable along these dimensions. Beliefs about the opponent's action choice are also similar throughout. However, these aggregate data mask important differences.

We now look at gender effects in Norway and India across the different treatments. The question we consider is whether participants condition their behavior based on the gender of their opponent. Accordingly, we measure a subject's propensity to select an action as a function of the subject's own gender and the opponent's gender. Since we are interested in understanding the extent to which gender effects may be driven by gender-dependent expectations, these are also included as an explanatory variable. Finally, within each country, our subjects differ in terms of socio-economic background and, in particular, education levels. As mentioned above, subjects with dissimilar educational outcomes are expected to differ in their beliefs, their gender attitudes, and their tolerance for inequality. To fully account for the interplay between education and gender effects, we estimate a model where the participant's education level interacts with the opponent's gender (when known to the participant) for both countries and for both men and women. The model to be estimated is accordingly

$$\Pr(Y_i = 1|X) = G\Bigg(\alpha_0 + \alpha_1 EDU_i + \alpha_2 KNOWNOPP_i + \alpha_3 FEMOPP_i + \alpha_4 KNOWNOPP_i \times EDU_i + \alpha_5 FEMOPP_i \times EDU_i + \alpha_6 EXPECT_i + \sum_{j=1}^{N} \beta_j X_{ji}\Bigg) \quad (1)$$

where $Y_i = 1$ if the participant chooses the hawkish option – the option which, if selected by both players, generates the highest monetary payoff for the participant – and $Y_i = 0$ otherwise. KNOWNOPP is a dummy returning one if the participant was informed of the opponent's gender, while FEMOPP returns one if the participant knew the opponent's gender and the opponent was a woman.[26] EDU is a dummy variable returning one if the participant is university educated.[27] The dummy EXPECT is used to control for the subject's beliefs about what the opponent is likely to play, and is equal to one if the participant expects the opponent

[26] It is important to stress that FEMOPP can only take a value of one if KNOWNOPP is one. If KNOWNOPP is zero, then FEMOPP is necessarily zero too.

[27] Since there are very few participants lacking a secondary school qualification, we pooled together those with secondary and primary school qualifications. This does not affect the key results.

to play hawkish. In addition to expectations about opponent's choice of action, we include the following set of further controls: (1) the participant's age; (2) whether the participant lives in a village, small city, or big city (LIVE); (3) whether the hawkish option was presented as Option A or Option B (ROLE) (more on this below).[28]

Since the interpretation of coefficients in non-linear models with interaction terms may be involved, in the main body, we consider a linear specification for G. The results do not depend on the specific functional form (Results for Probit can be found in the Supplementary Material).

A quick reference for the interpretation of the model's parameters is provided in Table 2.

Parameters	Propensity to play hawkish for...
α_0	low edu subj/gender unknown treat
$\alpha_0 + \alpha_1$	high edu subj/gender unknown treat
$\alpha_0 + \alpha_2$	low edu subj/male opponent
$\alpha_0 + \alpha_2 + \alpha_3$	low edu subj/fem opponent
$\alpha_0 + \alpha_1 + \alpha_2 + \alpha_4$	high edu subj/male opponent
$\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5$	high edu subj/fem opponent

TABLE 2. Description of coefficients

The gender effect coefficients are α_3 (the coefficient of FEMOPP) and α_5 (the coefficient of FEMOPP$\times$EDU), where

$$\begin{aligned} \alpha_3 : &\quad \text{(low edu/fem opp)} - \text{(low edu/male opp)} \\ \alpha_3 + \alpha_5 : &\ \text{(high edu/fem opp)} - \text{(high edu/male opp)} \end{aligned} \tag{2}$$

Results for India. Table 3 estimates model (1) for our Indian subject pool, including and excluding expectations.[29] Columns 1 and 2 consider the pooled sample without education interactions. In that case, α_3 – the coefficient of FEMOPP – is indistinguishable from zero. However, as previously discussed, pooled data may be misleading, since education has been shown to matter for gender attitudes. As can be seen from Columns 3 and 4, once education is interacted with FEMOPP, the coefficient becomes positive and significant. This signals the presence of a gender effect in the form of low education subjects conditioning their

[28] In the pooled regressions, we also included country fixed effect.

[29] Equivalent estimates for the pooled sample can be found in Table A.3 in the Appendix.

actions on the gender of the opponent and becoming more hawkish when facing a woman. A possible explanation for this finding is that it is *expectation-driven*: subjects may expect that the behavior of men and women differs, and that a woman opponent is less likely to play hawkish compared to a male. This hypothesis is however not corroborated by the data: if we compare the coefficient of FEMOPP in column 3 (where expectations are excluded) with that in column 4 (where expectations are included), we see that it hardly moves. This does not mean that expectations do not matter. Inspection of the coefficient of EXPECT, indicates that subjects are significantly less likely to play hawkish if they expect their opponent to behave hawkishly.

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish All	(4) Chose Hawkish All	(5) Chose Hawkish Female	(6) Chose Hawkish Female	(7) Chose Hawkish Male	(8) Chose Hawkish Male
EDU	0.035	0.039	0.049	0.074	0.076	0.090	0.020	0.052
	(0.040)	(0.037)	(0.065)	(0.060)	(0.096)	(0.090)	(0.089)	(0.082)
KNOWNOPP	-0.068*	-0.062	-0.113*	-0.091	-0.086	-0.064	-0.141	-0.116
	(0.040)	(0.038)	(0.066)	(0.059)	(0.095)	(0.085)	(0.091)	(0.084)
FEMOPP	0.044	0.035	0.159**	0.151***	0.208**	0.195**	0.132	0.129
	(0.039)	(0.037)	(0.062)	(0.058)	(0.092)	(0.084)	(0.084)	(0.080)
EXPECT Hawk		-0.340***		-0.341***		-0.348***		-0.333***
		(0.032)		(0.032)		(0.047)		(0.044)
KNOWNOPPXEDU			0.074	0.047	0.008	-0.003	0.145	0.101
			(0.083)	(0.078)	(0.120)	(0.111)	(0.116)	(0.110)
FEMOPPXEDU			-0.193**	-0.196***	-0.252**	-0.250**	-0.154	-0.161
			(0.080)	(0.075)	(0.116)	(0.108)	(0.111)	(0.105)
Constant	0.388***	0.560***	0.371***	0.530***	0.324***	0.603***	0.222**	0.449***
	(0.082)	(0.064)	(0.088)	(0.070)	(0.103)	(0.103)	(0.097)	(0.095)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.059	0.156	0.063	0.161	0.052	0.154	0.069	0.162
Observations	876	876	876	876	427	427	449	449

Notes: OLS estimates of the the baseline model. The outcome variable takes value one if the subject chose the hawkish action and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small city, big city), whether the hawkish option was denoted as Option A or B, and expectations about the opponent's strategy. * p < 0.10, ** p < 0.05, *** p < 0.01

TABLE 3. Propensity to choose the hawkish action- OLS estimates for India

This is in line with standard game-theoretic predictions. However, the key observation here is that including expectations has essentially *no effect* on the coefficients FEMOPP and FEMOPP×EDU (namely, α_3 and α_5) which are the coefficients of interest for gender effects. This suggests that the gender effect is *not* driven by subjects believing that male and female opponents are likely to behave differently.[30] We provide further support for this interpretation by showing that, if we regress player expectations (in the Gender Known treatment) against opponent gender, the expectations of those facing a man or a woman are essentially the same

[30]Note that our argument would continue to hold even in a world where subjects use their answers in the expectation question for hedging purposes. In such a world, the subjects' declared expectations would be the exact opposite of their true expectations – resulting in α_6, the coefficient of EXPECT, being statistically significant and *positive*. However, even in that case, if the observed gender effect was driven by the subjects' (true) expectations, including the expectation variable would eliminate the gender effect.

(see Figure A.1 in the Appendix). Overall, the evidence indicates that our subjects are not using gender as a coordination device in mixed matches. Rather, the effect appears to be more subtle, as further discussed in Section 4.1 below.

Another important observation is that the gender effect we identify in the data is driven by women with lower education. If anything, highly educated Indian women appear to be somewhat *less* hawkish when facing a woman, but this effect is not statistically significant. This can be seen from $\alpha_3 + \alpha_5$, namely the sum of the coefficients of FEMOPP and FEMOPP×EDU in columns 5 and 6, which is indistinguishable from zero.

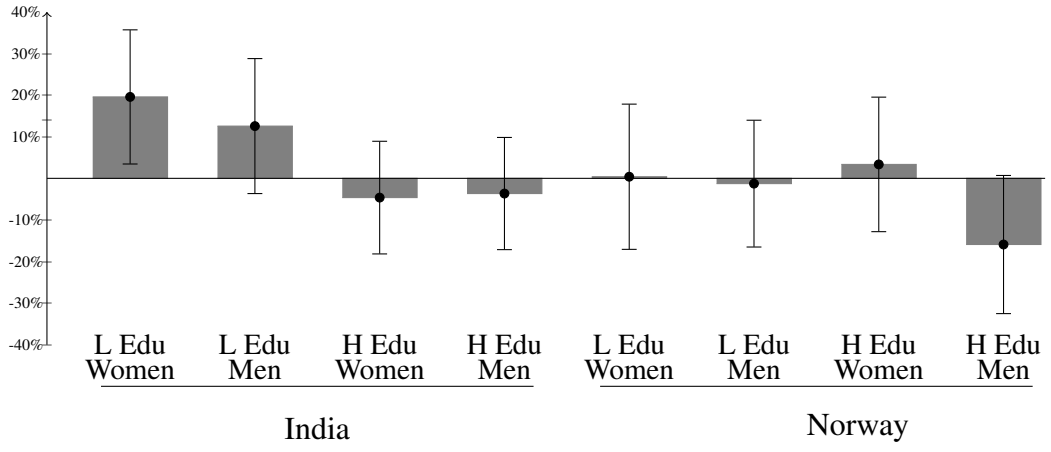


FIGURE 1. Difference in probability of playing hawkish when facing a female opponent compared to a male opponent.

The last two columns of the table consider male subjects. Highly educated men are unresponsive to their opponent's gender, similar to highly educated women. Men with lower education levels appear to somewhat condition their behavior on the gender of their opponent, but the coefficients of FEMOPP are smaller than for (lower-education) women and not statistically significant. This is in line with the observation that, in many experiments, men are less responsive to social cues when choosing their actions (Croson & Gneezy, 2009). To better visualize the results, Figure 1 reports for all subsamples the difference in the probability of choosing the hawkish option when facing a female opponent relative to a male opponent in the Gender Known treatment. This is positive and statistically significant for Indian women with lower education, who are about 20% more likely to play hawkish when facing a woman. For lower-education Indian men the gender effect is comparable in size albeit not significant,

while is entirely absent when considering highly educated subjects.[31]

The issue of how to interpret the role of education is a complex one, since educational attainments are clearly correlated with cultural traits and socio-economic status, so that no causal effect of education on behavior can be inferred. For instance, women with conservative gender role attitudes may be both more dovish when facing men *and* have a lower propensity to pursue higher education. While we cannot make claims about causality, our data allow us to explore whether the differences we observe between highly educated and less educated Indian subjects arise from different views about the role of men and women in the family and in society. In order to address this issue, we re-estimate the model by replacing education with various proxies for conservativeness of views on gender roles. In particular, we use answers to commonly used World Value Survey questions where we ask our participants to report the extent to which they agree with the following statements:

1. BEATING: 'It is fine for a husband to beat his wife if she behaves badly.'
2. JOBS: 'When jobs are scarce, men should have priority.'
3. INCOME: 'When a wife earns more than her husband, this causes problems within the household.'

Answers are recorded on a 1-5 scale, with 1 corresponding to 'strongly agree' and 5 to 'strongly disagree'. We also use an index which takes the average of the responses across the three questions. As one would expect, agreement with the statements tends to be negatively correlated with educational attainments.[32] However, we do not find that conservative views can explain the gender effect in our Indian data. When we replace education with views on gender roles derived from our survey questions, all the interactions with the opponent's gender return small and insignificant coefficients (see Table A.6 in the Supplementary Material). It thus seems unlikely that the observed differences between more and less educated subjects are driven by the fact that less educated subjects are more conservative when it comes to gender roles. The potential mechanisms for our results are further explored in Section 4.1 below.

[31] For completeness, in the Appendix we provide an analogous figure for expectations, rather than actions (See Figure A.1). The figure clearly shows that subjects by and large do not expect male and female opponents to behave very differently.

[32] The correlation between the answers and education levels persists also when controlling for age and place of residence.

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish All	(4) Chose Hawkish All	(5) Chose Hawkish Female	(6) Chose Hawkish Female	(7) Chose Hawkish Male	(8) Chose Hawkish Male
EDU	-0.057	-0.043	-0.070	-0.053	0.066	0.043	-0.224**	-0.176*
	(0.036)	(0.035)	(0.065)	(0.062)	(0.089)	(0.082)	(0.092)	(0.092)
KNOWNOPP	0.097**	0.108***	0.075	0.091	0.147*	0.176**	0.041	0.040
	(0.043)	(0.042)	(0.058)	(0.056)	(0.082)	(0.078)	(0.083)	(0.081)
FEMOPP	-0.028	-0.027	0.001	-0.004	-0.013	-0.043	-0.027	-0.003
	(0.042)	(0.040)	(0.058)	(0.057)	(0.087)	(0.087)	(0.080)	(0.076)
EXPECT Hawk		-0.301***		-0.301***		-0.325***		-0.261***
		(0.036)		(0.036)		(0.054)		(0.050)
KNOWNOPPXEDU			0.048	0.038	-0.191	-0.165	0.322***	0.274**
			(0.087)	(0.083)	(0.119)	(0.111)	(0.123)	(0.122)
FEMOPPXEDU			-0.059	-0.049	0.063	0.074	-0.166	-0.155
			(0.084)	(0.081)	(0.121)	(0.119)	(0.117)	(0.113)
Constant	0.358***	0.543***	0.362***	0.546***	0.176*	0.395***	0.546***	0.681***
	(0.062)	(0.062)	(0.067)	(0.067)	(0.093)	(0.095)	(0.095)	(0.094)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.022	0.097	0.020	0.095	0.015	0.093	0.061	0.119
Observations	802	802	802	802	403	403	399	399

Notes: OLS estimates of the the baseline model. The outcome variable takes value one if the subject chose the hawkish action and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small city, big city), whether the hawkish option was denoted as Option A or B, and expectations about the opponent's strategy. * p < 0.10, ** p < 0.05, *** p < 0.01

TABLE 4. Propensity to choose the hawkishish action - OLS estimates for Norway

Results for Norway. The picture for Norway differs substantially from India. Columns 1 to 4 in Table 4 indicate that, overall, Norwegian subjects do not condition their behavior on the gender of their opponent. This is also confirmed for female participants (columns 5 and 6 and 1). For male subjects (last two columns), however, we see that the coefficient of KNOWNOPP×EDU is positive and significant. This implies that, compared to the benchmark where the opponent's gender is unknown, highly educated men become more aggressive when they know they are facing another man, and this persists even when expectations are accounted for. On the other hand, when facing a woman, their behavior is the same as when they do not know the gender of their opponent.[33] A by-product of this is that, as highlighted in Figure 1, for highly educated men the probability of playing the hawkish action when facing a woman is smaller than when they face a male opponent. This suggests the presence of a gender effect, which, however, takes the opposite form of what we identify in India. Note that our findings for Norway differ substantially from those obtained by Holm (2000) in a comparably progressive country: Sweden. A possible explanation is that Holm's study was run more than 20 years ago, and therefore reflects a reality that may no longer be applicable. Indeed, his results are not confirmed by a later study by Boschini et al. (2014) of the Swedish population. Finally, our data on gender role attitudes indicate that, similar to the case of Indian subjects, the gender effect is not due to education acting as a proxy for traditionalism

[33] In the last two columns, the sum of the coefficients of KNOWNOPP×EDU and FEMOPP×EDU is not statistically different from zero.

of gender views in our Norvegian sample (see Table A.7).

4.1 Discussion

We now try to shed light on the possible mechanisms behind the gender effects identified above. As discussed in Section 3, there are a number of (non-mutually exclusive) reasons why participants may condition their behavior on the gender of the opponent:

1. **Coordination.** Participants use gender as a coordination device in mixed matches.
2. **Gender-specific preferences.** Men and women differ in their underlying propensity to play hawkish and, when playing in mixed matches, participants simply respond to these gender differences.
3. **Gender-dependent inequality aversion.** Inequality aversion depends on the opponent's gender.

1. **Coordination.** The first hypothesis considers the possibility that, in mixed matches, subjects use gender to solve the coordination problem they face in the game. For instance, if subjects expect their opponent to follow the 'man = dominant, woman =submissive' stereotype, then they find it optimal to follow it too. This mechanism crucially relies on gender acting as a catalyst that dictates how the subjects expect their opponent to play in mixed matches (an *expectation-driven* mechanism).

In our data, however, the gender effects persist even after expectations are controlled for. In fact, accounting for expectations hardly changes the coefficients of interest. This is not to say that participants do not recognise the need to identify focal points to help coordination. In our experiment, when a participant's hawkish option is denoted as A, Option A is also the opponent's dovish option, and this is common knowledge. Hence, to the extent that labelling a choice as A makes it more salient than labelling it as B, Option A provides a gender-independent focal point that participants may use to coordinate. Our data suggests that a good share of participants, in fact, did that: they tried to coordinate with their opponent by using the gender-independent coordination device of selecting the option denoted as A.[34] The key observation, though, is that gender does not appear to have been used as a

[34]The coefficient of the ROLE variable is significant at the 1% level for most subsamples (see full regression tables in the Appendix). Note that this gender-independent coordination device was not available in Holm (2000)'s study, since, in that study, options were not labelled.

coordination device in mixed matches.

2. **Gender-specific preferences.** Consider now the possibility that participants are simply responding to gender-related differences in preferences over hawkish or dovish behavior. For instance, if women are inherently more 'dovish' while men are inherently more 'hawkish', then it might be actually optimal to select the hawkish action with higher probability when facing a woman. To investigate this channel, we start by assessing whether these gender-related differences in the propensity to play hawkish exist in the first place. To isolate possible differences in intrinsic preferences from the effect of the opponent's gender, we restrict attention to the Gender Unknown treatment. The results in Table 5 show that there does not seem to be much difference in how men and women behave when they do not know the gender of their opponent. The only exception are highly educated Norwegian men, who have a lower propensity to play hawkish than similarly educated women, although the effect is weak. This suggests that men and women do not differ in their inherent propensity to play hawkish or dovish, and, hence, explanation 2. is unlikely to drive our findings. It is worth stressing that, similar to mechanism 1. discussed above, explanation 2. can also be ruled out by considering expectations. If women are inherently more 'dovish', this should be reflected in the participants' expectations, and any residual effect of the opponent's gender should disappear once we control for expectations. As already argued, however, this is not what we see in the data. All things considered, we do not find much support for the hypothesis that participants optimally respond to gender differences in inherent propensity to play hawkish.

3. **Gender-dependent inequality aversion.** The third possible explanation for the gender effects we observe in our data is that the subjects may experience different degrees of inequality aversion depending on the gender of their opponent. To explore this possibility, we divide the sample into participants who expect their opponent to choose the hawkish action and those who expect their opponent to play the dovish option. Suppose, for instance, that people are more averse to letting their opponent come out on top (disadvantageous inequality) when the opponent is a woman. Then, conditional on expecting the opponent to select the hawkish action, they should have a higher propensity to behave hawkishly when facing a woman. Looking at the subsample of subjects who expect their opponent to play the dovish action similarly allows us to shed light on whether the gender of the opponent affects the

subjects' aversion to advantageous inequality.

	ALL		INDIA		NORWAY	
	(1)	(2)	(3)	(4)	(5)	(6)
High EDU	-0.090	-0.053	-0.036	-0.011	-0.177*	-0.131
	(0.061)	(0.059)	(0.076)	(0.073)	(0.094)	(0.094)
Low EDU	-0.089	-0.061	-0.032	-0.040	-0.121	-0.077
	(0.065)	(0.061)	(0.102)	(0.092)	(0.084)	(0.081)
EXPECTATIONS	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.021	0.120	0.092	0.196	0.037	0.116
Obs	509	509	257	257	252	252

Notes: OLS estimates of gender differences in the propensity to play hawk. Robust standard errors in parentheses. The sample is restricted to the treatment in which the opponent's gender is unknown. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE 5. Gender differences in the propensity to choose the hawkish action in the Gender Unknown treatment (men-women)

This is what we do in Table 6. The table reports the effect of the opponent's gender on a subject's propensity to select the hawkish. Columns 1, 3 and 5 restrict attention to subjects who expect their opponent to behave hawkishly, while columns 2, 4 and 6 focus of subjects who expect their opponent to behave dovishly. As can be clearly seen from Table 6, our data suggest that aversion to advantageous inequality does not depend on the opponent's gender. However, aversion to disadvantageous inequality does: when subjects expect their opponent to select the hawkish action, they conform to the pattern illustrated in Table 1. This is particularly true of Indian women with lower education. In fact, in columns 1 and 3 the coefficient of FEMOPP is higher than in the baseline model of Table 2.[35] In contrast, when looking at columns 2 and 4, the coefficient of FEMOPP is lower than in Table 2, consistent with the idea that the gender effect may be driven by women who expect their opponent to behave hawkishly.

[35]It is important to stress that this does not imply that subjects are more hawkish against a woman opponent when they expect her to play hawkish. What it says is that the *difference* in the subjects' propensity to select the hawkish action when the opponent is a woman compared to when it is a man is greater for those subjects who expect their opponent to play hawkish.

	INDIA		INDIA: F		INDIA: M	
Expect opp to play:	Hawkish	Dovish	Hawkish	Dovish	Hawkish	Dovish
	(1)	(2)	(3)	(4)	(5)	(6)
EDU	0.155**	-0.119	0.167	-0.079	0.140	-0.144
	(0.076)	(0.097)	(0.113)	(0.155)	(0.103)	(0.128)
KNOWNOPP	-0.075	-0.094	-0.081	0.013	-0.056	-0.226
	(0.075)	(0.093)	(0.105)	(0.137)	(0.106)	(0.137)
FEMOPP	0.173**	0.088	0.241**	0.062	0.119	0.171
	(0.072)	(0.091)	(0.109)	(0.115)	(0.097)	(0.136)
KNOWNOPP×EDU	0.047	0.023	0.018	-0.094	0.070	0.166
	(0.097)	(0.126)	(0.138)	(0.183)	(0.136)	(0.179)
FEMOPP×EDU	-0.265***	-0.048	-0.325**	-0.051	-0.219*	-0.094
	(0.094)	(0.118)	(0.139)	(0.160)	(0.131)	(0.171)
Constant	0.287***	0.545***	0.353*	0.675***	0.221	0.430**
	(0.109)	(0.133)	(0.202)	(0.204)	(0.139)	(0.180)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.062	0.088	0.049	0.023	0.063	0.125
Observations	608	268	293	134	315	134

	NOR		NOR: F		NOR: M	
Expect opp to play:	Hawkish	Dovish	Hawkish	Dovish	Hawkish	Dovish
	(1)	(2)	(3)	(4)	(5)	(6)
EDU	-0.067	0.010	0.004	0.085	-0.149	-0.327*
	(0.074)	(0.116)	(0.098)	(0.150)	(0.110)	(0.168)
KNOWNOPP	0.103	0.090	0.208**	0.058	-0.008	0.116
	(0.069)	(0.097)	(0.090)	(0.166)	(0.108)	(0.124)
FEMOPP	-0.002	-0.039	-0.001	-0.136	-0.016	0.012
	(0.071)	(0.095)	(0.100)	(0.178)	(0.101)	(0.115)
KNOWNOPP×EDU	0.024	0.052	-0.186	0.053	0.318**	0.302
	(0.100)	(0.149)	(0.130)	(0.216)	(0.153)	(0.207)
FEMOPP×EDU	-0.046	-0.028	0.108	-0.095	-0.241*	0.046
	(0.099)	(0.144)	(0.137)	(0.238)	(0.144)	(0.186)
Constant	0.199**	0.747***	0.052	0.651***	0.361***	0.720***
	(0.080)	(0.117)	(0.105)	(0.213)	(0.126)	(0.149)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.028	-0.003	0.054	-0.048	0.036	0.069
Observations	564	238	300	103	264	135

Notes: Participants' choices conditional on whether they expect their opponent to play the hawkish or dovish action (OLS estimates). The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE 6. Propensity to choose the hawkish action– OLS estimates conditional on expected opponent's choice

While this should not be taken as fully conclusive evidence, it is highly suggestive that lower-education Indian women are more inclined to sacrifice monetary rewards to avoid finishing last when facing a female opponent relative to a male opponent. The gender effect we identify for Norway can be similarly rationalised by a mechanism of aversion to disadvantageous inequality which depends on the opponent's gender. Indeed, for highly educated Norwegian men, the sum of α_3 and α_5 is negative when the participant expects the opponent to play hawkish (column 5 in Table 6) and is essentially zero when dovish is instead anticipated (column 6).

5 Concluding Remarks

Our analysis prompts a number of observations. First, the gender effects we identify in the battle of the sexes differ substantially in the two countries we consider – Norway and India. This suggests that behavior in the game is informative about the environment the subjects operate in, and might potentially be used as a behavior-based measure of country-level gender inequality. It would be interesting to run the experiment in more countries in order to gather additional evidence and identify possible patterns.

Second, our investigation highlights the importance of considering subjects from a variety of backgrounds and, in particular, different levels of education, in order to gain an organic picture of gender effects in a society. Gender effects appear to vary substantially, depending on education. For instance, in our data university educated Indian subjects exhibit no gender effects at all, in contrast to what we find for subjects with lower levels of education. This suggests that different gender cultures may exist in different strata of the population.

Another intriguing finding of our analysis is that gender effects may also arise in the opposite direction to what one would normally expect, taking the form of subjects – in particular highly educated Norwegian men – being less hawkish when playing against a female opponent. A possible explanation could be that, in a country such as Norway which is highly attentive to gender issues, men may develop an aversion to behaving in a way that could potentially be interpreted as sexist. This could induce them to become less aggressive when interacting with a woman. The fact that this is observed only in highly educated subjects again points to the possible existence of potentially different sub-cultures in the population.

Finally, we find it very interesting that, in India, the gender effect appears to be stronger among female (less educated) subjects, who are more likely to adapt their behavior to the gender of their counterpart. Our analysis provides suggestive evidence that this is not necessarily because they expect a male opponent to be more 'hawkish', but rather it is due to female subjects being comparatively less averse to inequality when facing a man. There is a general point here, namely that what is considered 'fair' or acceptable in mixed gender interactions may differ across societies and socio-economic groups depending on prevailing social norms.[36] This may of course generate additional challenges when it comes

[36]For instance, Sardinha and Catalán (2018) report attitudes from 49 low- and middle-income countries about domestic violence. On average, 41% of women agree that a husband is sometimes justified in beating his wife.

to designing policies aimed at empowering disadvantaged groups. We believe that this is a very interesting area for future research.

A - Supplementary Tables

Country	F	M	FF	FM	MM	MF	Total
INDIA	125	132	141	161	156	161	874
NORWAY	130	122	124	149	128	149	802
Total	255	254	265	310	284	310	1676

TABLE A.1. Number of participants by country and treatment

The label F (resp, M) indicates women (resp., men) in the Gender Unknown treatment. The label Fx (where x is either F or M) indicates women in the Gender Known treatment who are matched with a subject of gender x. The label Mx is defined analogously.

	Overall	INDIA: F	INDIA: M	NORWAY: F	NORWAY: M
Correct recall COUNTRY	98%	99%	98%	97%	96%
Correct recall SEX	95%	91%	94%	99%	97%
Guess MALE	51%	59%	46%	54%	47%
PURPOSE OF THE STUDY					
Selfishness	53%	37%	51%	60%	63%
Competitiveness	33%	48%	39%	21%	20%
Gender	6%	6%	2%	9%	8%
Other	9%	9%	8%	1%	9%
N	1678	427	449	403	399

Notes: The table reports the shares of correctly recalled country and sex of Player 2. For the treatments in which the gender of the other player is not known, we asked participants to guess the gender of Player 2. Finally, we report the share of guesses about the study's objective.

TABLE A.2. Recall rates and study objective

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
INDIA	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
NORWAY	-0.055**	-0.048*	-0.069*	-0.047	-0.024	-0.035
	(0.027)	(0.026)	(0.038)	(0.036)	(0.040)	(0.038)
EDU	0.007	0.026	0.089	0.083	-0.069	-0.029
	(0.046)	(0.043)	(0.065)	(0.061)	(0.064)	(0.062)
KNOWNOPP	-0.008	0.011	0.040	0.063	-0.048	-0.037
	(0.044)	(0.041)	(0.062)	(0.058)	(0.062)	(0.059)
FEMOPP	0.069	0.063	0.098	0.076	0.041	0.054
	(0.043)	(0.040)	(0.063)	(0.061)	(0.059)	(0.055)
FEMOPP×EDU	-0.115**	-0.113**	-0.104	-0.098	-0.140*	-0.139*
	(0.057)	(0.055)	(0.083)	(0.080)	(0.080)	(0.076)
KNOWNOPP×EDU	0.034	0.014	-0.111	-0.104	0.184**	0.140*
	(0.060)	(0.057)	(0.084)	(0.079)	(0.085)	(0.082)
AGE	0.002**	0.002**	0.004***	0.004***	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.084**	-0.069*	-0.092*	-0.087*	-0.080	-0.055
	(0.037)	(0.036)	(0.054)	(0.052)	(0.050)	(0.049)
Big city	-0.044	-0.029	-0.055	-0.051	-0.034	-0.009
	(0.037)	(0.036)	(0.055)	(0.053)	(0.050)	(0.048)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.179***	0.171***	0.157***	0.146***	0.208***	0.203***
	(0.024)	(0.023)	(0.034)	(0.033)	(0.034)	(0.032)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.323***		-0.333***		-0.310***
		(0.024)		(0.035)		(0.033)
Constant	0.398***	0.594***	0.331***	0.558***	0.457***	0.618***
	(0.055)	(0.055)	(0.084)	(0.083)	(0.074)	(0.074)
R-squared	0.037	0.124	0.032	0.121	0.045	0.127
Observations	1678	1678	830	830	848	848

Notes: Full tables for the baseline model - All Sample. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B, expectations about the opponent's strategy, and country fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.3. Propensity to choose hawkish - pooled sample

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
EDU	0.049	0.074	0.076	0.090	0.020	0.052
	(0.065)	(0.060)	(0.096)	(0.090)	(0.089)	(0.082)
KNOWNOPP	-0.113*	-0.091	-0.086	-0.064	-0.141	-0.116
	(0.066)	(0.059)	(0.095)	(0.085)	(0.091)	(0.084)
FEMOPP	0.159**	0.151***	0.208**	0.195**	0.132	0.129
	(0.062)	(0.058)	(0.092)	(0.084)	(0.084)	(0.080)
FEMOPP×EDU	-0.193**	-0.196***	-0.252**	-0.250**	-0.154	-0.161
	(0.080)	(0.075)	(0.116)	(0.108)	(0.111)	(0.105)
KNOWNOPP×EDU	0.074	0.047	0.008	-0.003	0.145	0.101
	(0.083)	(0.078)	(0.120)	(0.111)	(0.116)	(0.110)
AGE	0.004***	0.003**	0.003	0.002	0.006**	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.100	-0.066	-0.231*	-0.138	-0.020	-0.015
	(0.072)	(0.071)	(0.119)	(0.115)	(0.091)	(0.092)
Big city	-0.097	-0.059	-0.168	-0.077	-0.075	-0.061
	(0.070)	(0.069)	(0.118)	(0.114)	(0.086)	(0.087)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.232***	0.223***	0.209***	0.194***	0.254***	0.250***
	(0.033)	(0.031)	(0.047)	(0.045)	(0.046)	(0.044)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.341***		-0.348***		-0.333***
		(0.032)		(0.047)		(0.044)
Constant	0.371***	0.589***	0.492***	0.680***	0.297***	0.511***
	(0.088)	(0.088)	(0.145)	(0.140)	(0.112)	(0.115)
R-squared	0.063	0.161	0.052	0.154	0.069	0.162
Observations	876	876	427	427	449	449

Notes: Full tables for the baseline model - India. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B, and expectations about the opponent's strategy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.4. Propensity to choose hawkish - Indian sample

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
EDU	-0.070	-0.053	0.066	0.043	-0.224**	-0.176*
	(0.065)	(0.062)	(0.089)	(0.082)	(0.092)	(0.092)
KNOWNOPP	0.075	0.091	0.147*	0.176**	0.041	0.040
	(0.058)	(0.056)	(0.082)	(0.078)	(0.083)	(0.081)
FEMOPP	0.001	-0.004	-0.013	-0.043	-0.027	-0.003
	(0.058)	(0.057)	(0.087)	(0.087)	(0.080)	(0.076)
FEMOPP×EDU	-0.059	-0.049	0.063	0.074	-0.166	-0.155
	(0.084)	(0.081)	(0.121)	(0.119)	(0.117)	(0.113)
KNOWNOPP×EDU	0.048	0.038	-0.191	-0.165	0.322***	0.274**
	(0.087)	(0.083)	(0.119)	(0.111)	(0.123)	(0.122)
AGE	0.001	0.002	0.005**	0.005***	-0.002	-0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.079*	-0.069	-0.031	-0.037	-0.130**	-0.103*
	(0.044)	(0.043)	(0.064)	(0.062)	(0.061)	(0.060)
Big city	-0.028	-0.017	-0.033	-0.043	-0.004	0.024
	(0.045)	(0.043)	(0.065)	(0.063)	(0.063)	(0.060)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.118***	0.111***	0.088*	0.078	0.163***	0.156***
	(0.035)	(0.034)	(0.050)	(0.049)	(0.049)	(0.048)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.301***		-0.325***		-0.261***
		(0.036)		(0.054)		(0.050)
Constant	0.390***	0.563***	0.209**	0.438***	0.550***	0.657***
	(0.070)	(0.070)	(0.101)	(0.104)	(0.097)	(0.096)
R-squared	0.020	0.095	0.015	0.093	0.061	0.119
Observations	802	802	403	403	399	399

Notes: Full tables for the baseline model - Norway. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B , and expectations about the opponent's strategy. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE A.5. Propensity to choose hawkish - Norwegian sample

Dep var: propensity to choose hawkish						
INDIA	Women			Men		
	beating	jobs	income	beating	jobs	income
	(1)	(2)	(3)	(4)	(5)	(6)
KNOWNOPP	-0.212	-0.131	-0.066	-0.038	-0.125	-0.015
	(0.171)	(0.158)	(0.153)	(0.202)	(0.167)	(0.167)
FEMOPP	-0.152	-0.055	0.021	-0.165	0.155	-0.100
	(0.169)	(0.153)	(0.147)	(0.186)	(0.161)	(0.162)
BEATING	-0.089***			-0.024		
	(0.033)			(0.035)		
KNOWNOPP×BEATING	0.030			-0.003		
	(0.040)			(0.046)		
FEMOPP×BEATING	0.049			0.049		
	(0.039)			(0.043)		
JOBS		-0.044			-0.015	
		(0.029)			(0.030)	
KNOWNOPP×JOBS		0.014			0.018	
		(0.038)			(0.041)	
FEMOPP×JOBS		0.025			-0.029	
		(0.036)			(0.039)	
INCOME			-0.021			-0.029
			(0.033)			(0.034)
KNOWNOPP×INCOME			-0.006			-0.012
			(0.043)			(0.045)
FEMOPP×INCOME			0.010			0.041
			(0.042)			(0.043)
Constant	0.996***	0.750***	0.639***	0.374**	0.333*	0.388**
	(0.201)	(0.187)	(0.189)	(0.183)	(0.173)	(0.167)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.061	0.045	0.043	0.064	0.063	0.066
Observations	427	427	427	449	449	449

Notes: OLS estimates of propensity to choose hawkish conditional on conservative/liberal gender attitudes in India. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. BEATING: Agreement with the statement 'It is fine for a husband to beat his wife if she behaves badly'. JOBS: Agreement with 'When jobs are scarce, men should have priority'. INCOME: Agreement with 'When a wife earns more than her husband, this causes problems within the household'. Answers range from 1 to 5, where 1 = strongly agree and 5 = strongly disagree. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE A.6. Role of conservative/liberal gender attitudes – India

Dep var: propensity to choose hawkish

NORWAY	Women			Men		
	beating	jobs	income	beating	jobs	income
	(1)	(2)	(3)	(4)	(5)	(6)
KNOWNOPP	0.475	0.398	-0.049	0.784**	0.091	0.348
	(0.580)	(0.417)	(0.232)	(0.329)	(0.318)	(0.239)
FEMOPP	-0.101	-0.469	0.153	-0.253	-0.138	-0.279
	(0.251)	(0.433)	(0.242)	(0.489)	(0.376)	(0.249)
BEATING	-0.073			0.098**		
	(0.112)			(0.048)		
KNOWNOPP×BEATING	-0.085			-0.126*		
	(0.118)			(0.069)		
FEMOPP×BEATING	0.023			0.031		
	(0.053)			(0.100)		
JOBS		-0.042			-0.025	
		(0.077)			(0.052)	
KNOWNOPPXJOBS		-0.072			0.020	
		(0.087)			(0.071)	
FEMOPP×JOBS		0.102			0.007	
		(0.090)			(0.080)	
INCOME			-0.016			0.038
			(0.042)			(0.042)
KNOWNOPP×INCOME			0.027			-0.043
			(0.057)			(0.059)
FEMOPP×INCOME			-0.034			0.041
			(0.059)			(0.059)
Constant	0.614	0.459	0.313	0.022	0.598**	0.332*
	(0.563)	(0.381)	(0.194)	(0.234)	(0.249)	(0.189)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.022	0.016	0.007	0.049	0.043	0.046
Observations	403	403	403	399	399	399

Notes: OLS estimates of propensity to choose hawkish conditional on conservative/liberal gender attitudes in Norway. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. BEATING: Agreement with the statement 'It is fine for a husband to beat his wife if she behaves badly'. JOBS: Agreement with 'When jobs are scarce, men should have priority'. INCOME: Agreement with 'When a wife earns more than her husband, this causes problems within the household'. Answers range from 1 to 5, where 1 = strongly agree and 5 = strongly disagree. Controls are the same as in the baseline model (see Table **3**). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.7. Role of conservative/liberal gender attitudes – Norway

	INDIA				NORWAY			
	L EDU Women	H EDU Women	L EDU Men	H EDU Men	L EDU Women	H EDU Women	L EDU Men	H EDU Men
FEMOPP	0.196**	-0.046	0.126	-0.037	0.004	0.033	-0.013	-0.159*
	(0.084)	(0.070)	(0.082)	(0.069)	(0.088)	(0.082)	(0.076)	(0.085)
AGE	-0.004	-0.002	0.008**	-0.000	0.002	0.005	-0.001	0.001
	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
LIVE								
Village	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Small town	0.082	-0.444*	0.041	-0.093	-0.063	-0.088	-0.186**	-0.136
	(0.126)	(0.265)	(0.137)	(0.179)	(0.103)	(0.111)	(0.091)	(0.108)
Big city	0.041	-0.222	0.002	-0.054	-0.023	-0.121	0.020	0.002
	(0.126)	(0.249)	(0.136)	(0.156)	(0.111)	(0.110)	(0.105)	(0.109)
ROLE								
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.154*	0.241***	0.165**	0.318***	0.250***	0.101	0.189**	0.191**
	(0.082)	(0.069)	(0.082)	(0.069)	(0.087)	(0.082)	(0.078)	(0.086)
EXPECTATIONS								
EXPECT Dove	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
EXPECT Hawk	-0.545***	-0.264***	-0.333***	-0.288***	-0.151	-0.354***	-0.320***	-0.280***
	(0.086)	(0.071)	(0.088)	(0.069)	(0.098)	(0.097)	(0.078)	(0.090)
Constant	0.778***	0.907***	0.338*	0.671***	0.481***	0.575***	0.764***	0.703***
	(0.159)	(0.267)	(0.185)	(0.195)	(0.162)	(0.189)	(0.134)	(0.189)
R-squared	0.256	0.143	0.106	0.155	0.045	0.081	0.155	0.128
Observations	115	187	138	179	134	139	148	129

Notes: Change in probability of playing hawk when facing a female opponent compared to a male opponent. For each subgroup, we estimated the model $y_i = \gamma_0 + \gamma_1 FEMOPP_i + \sum_{j=1}^{n} \beta_j X_{ij} + \varepsilon_i$. The figure shows the coefficients for FEMOPP and their 95% confidence intervals for each subgroup. Controls are the same as in the baseline model (including expectations). Robust standard errors in parentheses. The sample was restricted to the treatments where the opponent's gender is known. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE A.8. Regression output for Figure 1

A.1 Differences in expectations according to the opponent's gender (female – male opponent)

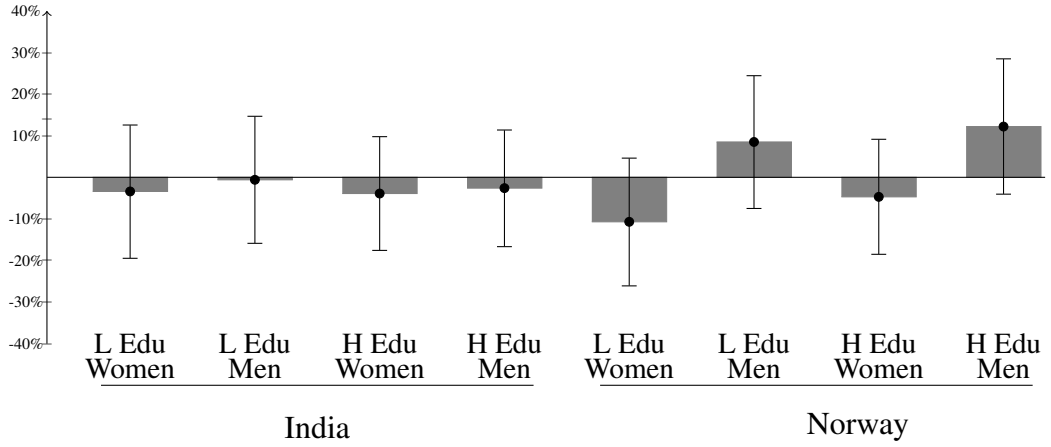


FIGURE A.1. Difference in the probability of saying that the opponent will play hawkish when facing a female opponent compared to a male opponent.

B - Probit estimates

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
INDIA	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
NORWAY	-0.143**	-0.139*	-0.178*	-0.141	-0.062	-0.104
	(0.070)	(0.072)	(0.098)	(0.100)	(0.104)	(0.108)
EDU	0.017	0.073	0.231	0.234	-0.180	-0.079
	(0.118)	(0.121)	(0.166)	(0.170)	(0.167)	(0.174)
KNOWNOPP	-0.022	0.036	0.105	0.193	-0.126	-0.105
	(0.113)	(0.117)	(0.160)	(0.164)	(0.162)	(0.166)
FEMOPP	0.178	0.177	0.254	0.206	0.106	0.159
	(0.110)	(0.113)	(0.163)	(0.169)	(0.153)	(0.154)
FEMOPP×EDU	-0.299**	-0.316**	-0.269	-0.270	-0.368*	-0.392*
	(0.149)	(0.153)	(0.215)	(0.223)	(0.210)	(0.215)
KNOWNOPP×EDU	0.088	0.038	-0.290	-0.302	0.487**	0.392*
	(0.154)	(0.159)	(0.216)	(0.221)	(0.221)	(0.230)
AGE	0.006**	0.007***	0.011***	0.011***	0.001	0.003
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.218**	-0.188*	-0.241*	-0.243*	-0.210	-0.144
	(0.096)	(0.099)	(0.142)	(0.147)	(0.131)	(0.136)
Big city	-0.114	-0.075	-0.145	-0.144	-0.091	-0.012
	(0.096)	(0.099)	(0.142)	(0.148)	(0.131)	(0.135)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.455***	0.472***	0.404***	0.407***	0.534***	0.561***
	(0.062)	(0.064)	(0.089)	(0.092)	(0.088)	(0.091)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.883***		-0.909***		-0.854***
		(0.072)		(0.105)		(0.099)
Constant	-0.260*	0.251	-0.433**	0.151	-0.106	0.321
	(0.142)	(0.153)	(0.217)	(0.231)	(0.191)	(0.207)
R-squared	0.031	0.099	0.032	0.102	0.042	0.106
LR χ^2	70.699	215.941	36.057	109.082	48.431	116.905
Observations	1678	1678	830	830	848	848

Notes: Full tables for the baseline model - Pooled sample. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B, expectations about the opponent's strategy, and country fixed effects. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.1. Probit estimates of the baseline model – Pooled sample

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
EDU	0.128	0.197	0.202	0.260	0.050	0.123
	(0.172)	(0.177)	(0.252)	(0.259)	(0.240)	(0.247)
KNOWNOPP	-0.300*	-0.257	-0.225	-0.168	-0.374	-0.335
	(0.175)	(0.177)	(0.251)	(0.251)	(0.246)	(0.251)
FEMOPP	0.419**	0.435***	0.553**	0.579**	0.351	0.364
	(0.165)	(0.169)	(0.247)	(0.250)	(0.226)	(0.233)
FEMOPP×EDU	-0.512**	-0.566***	-0.670**	-0.743**	-0.415	-0.452
	(0.212)	(0.219)	(0.309)	(0.317)	(0.296)	(0.306)
KNOWNOPP×EDU	0.197	0.134	0.021	-0.013	0.382	0.286
	(0.221)	(0.228)	(0.316)	(0.323)	(0.312)	(0.323)
AGE	0.011**	0.011**	0.008	0.005	0.016**	0.018**
	(0.004)	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.262	-0.158	-0.630*	-0.405	-0.051	-0.018
	(0.195)	(0.207)	(0.342)	(0.360)	(0.243)	(0.260)
Big city	-0.252	-0.139	-0.464	-0.241	-0.198	-0.140
	(0.188)	(0.200)	(0.337)	(0.355)	(0.231)	(0.248)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.601***	0.640***	0.543***	0.557***	0.664***	0.721***
	(0.087)	(0.090)	(0.125)	(0.129)	(0.123)	(0.128)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.977***		-0.985***		-0.979***
		(0.101)		(0.145)		(0.144)
Constant	-0.341	0.215	0.000	0.499	-0.553*	-0.004
	(0.235)	(0.259)	(0.398)	(0.431)	(0.306)	(0.337)
R-squared	0.054	0.134	0.054	0.136	0.066	0.143
LR χ^2	62.494	144.839	29.540	73.517	39.851	79.766
Observations	876	876	427	427	449	449

Notes: Full tables for the baseline model - India. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B, and expectations about the opponent's strategy. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.2. Probit estimates of the baseline model – India

	(1) Chose Hawkish All	(2) Chose Hawkish All	(3) Chose Hawkish Female	(4) Chose Hawkish Female	(5) Chose Hawkish Male	(6) Chose Hawkish Male
EDU	-0.177	-0.140	0.175	0.124	-0.597**	-0.473*
	(0.168)	(0.171)	(0.229)	(0.232)	(0.255)	(0.263)
KNOWNOPP	0.192	0.251	0.381*	0.492**	0.112	0.112
	(0.149)	(0.154)	(0.212)	(0.219)	(0.215)	(0.219)
FEMOPP	0.002	-0.008	-0.035	-0.127	-0.076	-0.002
	(0.149)	(0.154)	(0.223)	(0.233)	(0.210)	(0.210)
FEMOPP×EDU	-0.150	-0.138	0.164	0.204	-0.434	-0.436
	(0.215)	(0.220)	(0.309)	(0.320)	(0.308)	(0.314)
KNOWNOPP×EDU	0.121	0.102	-0.500	-0.455	0.863**	0.753**
	(0.224)	(0.228)	(0.307)	(0.310)	(0.336)	(0.347)
AGE	0.003	0.005	0.013***	0.015***	-0.006	-0.003
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
LIVE						
Village	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Small town	-0.201*	-0.190	-0.083	-0.105	-0.342**	-0.277*
	(0.113)	(0.116)	(0.164)	(0.168)	(0.163)	(0.165)
Big city	-0.073	-0.047	-0.089	-0.117	-0.012	0.066
	(0.115)	(0.117)	(0.165)	(0.169)	(0.167)	(0.170)
ROLE						
Hawk Option B	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Hawk Option A	0.302***	0.300***	0.227*	0.211	0.431***	0.431***
	(0.090)	(0.092)	(0.129)	(0.132)	(0.130)	(0.133)
EXPECTATIONS						
EXPECT Dove		0.000		0.000		0.000
		(.)		(.)		(.)
EXPECT Hawk		-0.800***		-0.869***		-0.708***
		(0.103)		(0.157)		(0.141)
Constant	-0.278	0.164	-0.750***	-0.187	0.137	0.430
	(0.177)	(0.189)	(0.265)	(0.285)	(0.249)	(0.264)
R-squared	0.023	0.080	0.027	0.087	0.062	0.107
LR χ^2	24.480	82.317	15.071	43.742	30.536	49.307
Observations	802	802	403	403	399	399

Notes: Full tables for the baseline model - Norway. The outcome variable takes value one if the subject chose the hawkish option and zero otherwise. Robust standard errors in parentheses. Controls are: age, place of residence (village, small town, big city), whether the hawkish option was denoted as Option A or B, and expectations about the opponent's strategy.. * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.3. Probit estimates of the baseline model – Norway

Conservative/liberal attitudes

Dep var: propensity to choose hawkish

INDIA	Women			Men		
	beating	jobs	income	beating	jobs	income
	(1)	(2)	(3)	(4)	(5)	(6)
KNOWNOPP	-0.632	-0.372	-0.169	-0.082	-0.352	-0.038
	(0.553)	(0.460)	(0.405)	(0.581)	(0.480)	(0.467)
FEMOPP	-0.436	-0.154	0.055	-0.471	0.417	-0.274
	(0.463)	(0.413)	(0.387)	(0.518)	(0.441)	(0.439)
BEATING	-0.258**			-0.069		
	(0.104)			(0.098)		
KNOWNOPP×BEATING	0.095			-0.013		
	(0.124)			(0.132)		
FEMOPP×BEATING	0.137			0.136		
	(0.106)			(0.120)		
JOBS		-0.125			-0.046	
		(0.085)			(0.088)	
KNOWNOPP×JOBS		0.042			0.052	
		(0.107)			(0.115)	
FEMOPP×JOBS		0.069			-0.078	
		(0.097)			(0.106)	
INCOME			-0.056			-0.079
			(0.088)			(0.095)
KNOWNOPP×INCOME			-0.017			-0.033
			(0.114)			(0.125)
FEMOPP×INCOME			0.029			0.110
			(0.111)			(0.117)
Constant	1.445**	0.720	0.380	-0.332	-0.437	-0.308
	(0.614)	(0.533)	(0.507)	(0.514)	(0.483)	(0.459)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.061	0.049	0.046	0.062	0.062	0.064
LR χ^2	34.437	28.279	25.793	39.364	39.223	39.457
Observations	427	427	427	449	449	449

Notes: Probit estimates of propensity to play hawk conditional on conservative/liberal gender attitudes in India. Robust standard errors in parentheses. BEATING: Agreement with the statement 'It is fine for a husband to beat his wife if she behaves badly'. JOBS: Agreement with 'When jobs are scarce, men should have priority'. INCOME: Agreement with 'When a wife earns more than her husband, this causes problems within the household'. Answers range from 1 to 5, where 1 = strongly agree and 5 = strongly disagree. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.4. Role of conservative/liberal gender attitudes – India

Dep var: propensity to choose hawkish

NORWAY	Women			Men		
	beating	jobs	income	beating	jobs	income
	(1)	(2)	(3)	(4)	(5)	(6)
KNOWNOPP	3.251	1.265	-0.131	2.327**	0.245	0.961
	(3.003)	(1.173)	(0.587)	(1.157)	(0.822)	(0.642)
FEMOPP	-1.728	-1.431	0.400	-0.707	-0.370	-0.774
	(2.889)	(1.260)	(0.621)	(1.349)	(0.989)	(0.660)
BEATING	-0.185			0.303*		
	(0.283)			(0.181)		
KNOWNOPP×BEATING	-0.627			-0.385		
	(0.604)			(0.239)		
FEMOPP×BEATING	0.353			0.089		
	(0.581)			(0.276)		
JOBS		-0.108			-0.067	
		(0.189)			(0.133)	
KNOWNOPP×JOBS		-0.235			0.052	
		(0.244)			(0.182)	
FEMOPP×JOBS		0.308			0.020	
		(0.261)			(0.211)	
INCOME			-0.043			0.104
			(0.107)			(0.112)
KNOWNOPP×INCOME			0.070			-0.122
			(0.145)			(0.157)
FEMOPP×INCOME			-0.089			0.117
			(0.151)			(0.156)
Constant	0.277	-0.110	-0.472	-1.476*	0.271	-0.462
	(1.414)	(0.937)	(0.491)	(0.895)	(0.631)	(0.514)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.035	0.029	0.022	0.053	0.048	0.050
LR χ^2	18.715	17.223	12.122	29.344	25.954	27.720
Observations	403	403	403	399	399	399

Notes: Probit estimates of propensity to play hawk conditional on conservative/liberal gender attitudes in Norway. Robust standard errors in parentheses. BEATING: Agreement with the statement 'It is fine for a husband to beat his wife if she behaves badly'. JOBS: Agreement with 'When jobs are scarce, men should have priority'. INCOME: Agreement with 'When a wife earns more than her husband, this causes problems within the household'. Answers range from 1 to 5, where 1 = strongly agree and 5 = strongly disagree. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.5. Role of conservative/liberal gender attitudes – Norway

Subsample splits according to expectations

	INDIA		INDIA: F		INDIA: M	
Expect opp play:	Hawk	Dove	Hawk	Dove	Hawk	Dove
	(1)	(2)	(3)	(4)	(5)	(6)
EDU	0.422**	-0.500	0.449	-0.232	0.389	-0.770
	(0.208)	(0.395)	(0.307)	(0.531)	(0.287)	(0.625)
KNOWNOPP	-0.210	-0.374	-0.244	0.116	-0.145	-0.984*
	(0.219)	(0.374)	(0.318)	(0.515)	(0.305)	(0.591)
FEMOPP	0.481**	0.313	0.688**	0.264	0.325	0.615
	(0.204)	(0.328)	(0.309)	(0.533)	(0.274)	(0.442)
KNOWNOPP×EDU	0.135	0.176	0.081	-0.350	0.180	0.822
	(0.271)	(0.472)	(0.392)	(0.644)	(0.378)	(0.725)
FEMOPP×EDU	-0.725***	-0.229	-0.913**	-0.264	-0.594*	-0.403
	(0.261)	(0.413)	(0.384)	(0.627)	(0.361)	(0.591)
Constant	-0.580*	-0.033	-0.397	0.352	-0.766**	-0.351
	(0.301)	(0.456)	(0.555)	(0.699)	(0.386)	(0.671)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.057	0.114	0.060	0.086	0.068	0.183
LR χ^2	43.010	32.972	22.094	14.694	26.311	25.076
Observations	608	268	293	134	315	134

	NOR		NOR: F		NOR: M	
Expect opp play:	Hawk	Dove	Hawk	Dove	Hawk	Dove
	(1)	(2)	(3)	(4)	(5)	(6)
EDU	-0.190	0.046	0.018	0.263	-0.402	-0.922*
	(0.208)	(0.311)	(0.291)	(0.448)	(0.300)	(0.502)
KNOWNOPP	0.272	0.270	0.577**	0.172	-0.023	0.342
	(0.184)	(0.278)	(0.249)	(0.469)	(0.280)	(0.358)
FEMOPP	-0.004	-0.110	-0.008	-0.372	-0.041	0.095
	(0.182)	(0.283)	(0.260)	(0.479)	(0.266)	(0.365)
KNOWNOPP×EDU	0.076	0.149	-0.511	0.202	0.849**	0.863
	(0.273)	(0.433)	(0.371)	(0.699)	(0.412)	(0.615)
FEMOPP×EDU	-0.124	-0.115	0.296	-0.361	-0.640*	0.049
	(0.257)	(0.434)	(0.360)	(0.699)	(0.384)	(0.592)
Constant	-0.793***	0.663*	-1.221***	0.378	-0.368	0.609
	(0.220)	(0.339)	(0.307)	(0.601)	(0.333)	(0.435)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.033	0.028	0.063	0.036	0.052	0.106
LR χ^2	23.723	8.647	25.099	4.168	16.815	18.222
Observations	564	238	300	103	264	135

Notes: Participants' choices conditional on whether they expect their opponent to play hawk or dove (Probit estimates). Robust standard errors in parentheses. Controls are the same as in the baseline model (see Table **3**). * p < 0.10, ** p < 0.05, *** p < 0.01.

TABLE B.6. Propensity to choose hawkish – Probit estimates conditional on expected opponent's choice

Chapter V

Conclusion

Each of the chapters in the thesis contributes to a different literature. Chapter 2 contributes to two streams of previous research: the one examining self-selection into different incentive types and the one studying the determinants of self-selection into the mission sector. In recent years, there has been a debate about the use of high-powered incentives as a mean to increase productivity in the mission/public sector. However, apart from productivity concerns, it is also important to keep in mind that prosociality is another key factor given the type of services provided within the sector. The evidence accumulated in the literature suggests that interacting prosocial motives and financial incentives might affect intrinsic motivation (for a summary, see Gneezy et al., 2011). Therefore, it is reasonable to believe that the use of performance-based incentives could affect the average prosociality of the applicant's pool negatively.

In particular, my findings point to the fact that care is needed when choosing which incentive scheme to offer. From the perspective of the mission oriented employer, it is fundamental to consider what other mission-driven organisations are offering as it could impact the type of subjects applying for the job. When all mission sector organisations offer a high-powered incentive, altruism ceases to be a significant predictor of the choice of the mission sector. Whereas when some of them provide a low-powered incentive, the organisations working in the mission sector could benefit from introducing a high-powered incentive, as it would enable them to attract more productive and prosocial individuals. Also, providing high-powered incentives might not be as beneficial as initially thought, especially if one considers that overall they were not able to outshine other incentive structures both in terms of donations raised and selection. In future extensions, it would be interesting to observe how the same question performs in a more natural setting, such as an actual mission-driven organisation. Similar field experiments looking at the effect of higher financial incentives already exist (Deserranno, 2019; Ashraf et al., 2020); however, to the best of my knowledge, there is no experiment examining the effect of high-powered incentives on selection into the mission sector.

The results of Chapter 3 indicate that individual preferences for competition extend to choosing on behalf of others, however, not in the anticipated direction. We find that more competitive and risk-loving individuals are more likely to send their agents into the tournament, but we also find no gender gap. That is, the decision to make others compete does not seem to be determined by the gender of the principal, or by that of the agent. We also

find mixed effects on efficiency. Whereas some measures suggest an efficiency loss when principals take the tournament entry decision, others suggest no efficiency loss compared to the case in which subjects take the decision themselves.

Our results strengthen the push for roles within organisations that could advise or, even, take this type of decision on behalf of the subjects. By no means, however, these results are to be interpreted as a motivation to strip individuals of their freedom of choice. They represent an extreme case that was designed expressively to answer our research question. When it comes to future research, some suggestions include a setting where subjects can choose whether or not to delegate the decision to someone else. Perhaps it would be interesting to see if a gender gap arises in such a context. Also, a setting where the earnings of the person deciding and that entering the competition are not aligned could be useful.

Chapter 4 examines the effect of gender labels in the context of a battle of the sexes game. The experiment finds evidence of the importance of social norms and their interaction with the choice of strategy between cultures. In particular, in India, subjects tend to be more likely to choose the aggressive strategy when facing a woman. The driver of this effect are mainly low-education women. However, in Norway, we find the opposite result, highly educated men are more likely to choose the aggressive strategy when facing other men. Our evidence points to the fact that the mechanisms behind these effects are related to inequality aversion, where subjects are more accepting of inequality in a way that is influenced by gender and culture.

These results have implications relating to policy. In particular, they suggest that when implementing empowerment programs in very gender unequal societies, it is important to consider entitlements. That is, the aim should be to reinforce female entitlement first, for example, with the use of female role models or by trying to change the way in which women perceive their roles (see Chattopadhyay & Duflo, 2004; Beaman et al., 2012) and then proceed to implement programs aiming at broadening the options available to women. Future research should explore in more detail the interaction between disadvantageous inequality and gender norms, focusing in particular on their influence on real-life outcomes such as labour force participation or household bargaining situations.

References

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, *115*(3), 715–753.

Aksoy, B., Chadd, I., & Koh, B. H. (2021). (anticipated) discrimination against sexual minorities in prosocial domains. *Available at SSRN 3903425*.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., & Tungodden, B. (2016a). What explains the gender gap in college track dropout? Experimental and administrative evidence. *American Economic Review*, *106*(5), 296–302.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., & Tungodden, B. (2016b). Willingness to compete: Family matters. *Management Science*, *62*(8), 2149–2162.

Almås, I., Cappelen, A. W., & Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, *128*(5), 1753–1788.

Ambuehl, S., Bernheim, B. D., & Ockenfels, A. (2021). Projective paternalism. *American Economic Review*, *111*(3).

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, *95*(4), 1438–1443.

Apicella, C. L., Demiral, E. E., & Mollerstrom, J. (2017). No gender difference in willingness to compete when competing against self. *American Economic Review*, *107*(5), 136–40.

Ashraf, N., Bandiera, O., Davenport, E., & Lee, S. S. (2020). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services. *American Economic Review*, *110*(5), 1355–94.

Atkin, D. (2013). Trade, tastes, and nutrition in india. *American Economic Review*, *103*(5), 1629–63.

Ayres, I., & Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *American Economic Review*, 304–321.

Balafoutas, L., Fornwagner, H., & Sutter, M. (2018). Closing the gender gap in competitiveness through priming. *Nature Communications*, *9*(1), 1–6.

Balafoutas, L., & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, *335*(6068), 579–582.

Balafoutas, L., & Sutter, M. (2019). How uncertainty and ambiguity in tournaments affect gender differences in competitive behavior. *European Economic Review*, *118*, 1–13.

Baldiga, N. R., & Coffman, K. B. (2018). Laboratory evidence on the effects of sponsorship on the competitive preferences of men and women. *Management Science*, *64*(2), 888–901.

Banuri, S., & Keefer, P. (2013). Intrinsic motivation, effort and the call to public service. *World Bank Policy research working paper*(6729).

Banuri, S., & Keefer, P. (2016). Pro-social motivation, effort and the call to public service. *European Economic Review*, *83*, 139–164.

Barigozzi, F., Burani, N., & Raggi, D. (2018). Productivity crowding-out in labor markets with motivated workers. *Journal of Economic Behavior & Organization*, *151*, 199–218.

Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *Science*, *335*(6068), 582–586.

Becker, A., Enke, B., & Falk, A. (2020). Ancient origins of the global variation in economic preferences. *AEA Papers and Proceedings*, *110*, 319–23.

Bellé, N., & Cantarelli, P. (2015). Monetary incentives, motivation, and job effort in the public sector: An experimental study with italian government executives. *Review of Public Personnel Administration*, *35*(2), 99–123.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, *96*(5), 1652–1678.

Ben-Ner, A., Kong, F., & Putterman, L. (2004). Share and share alike? intelligence, socialization, personality, and gender-pairing as determinants of giving. *Journal of Economic Psychology*, *25*(5), 581–589.

Berge, L. I. O., Bjorvatn, K., Pires, A. J. G., & Tungodden, B. (2015). Competitive in the lab, successful in the field? *Journal of Economic Behavior & Organization*, *118*, 303–317.

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, *95*(3), 616–636.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, philosophy & economics*, *9*(3), 297–313.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., & Borghi, J. (2015). Effect of paying for performance on utilisation, quality, and user costs of health services in tanzania: a controlled before and after study. *PloS one*, *10*(8), e0135013.

Black, S. E., & Strahan, P. E. (2001). The division of spoils: Rent-sharing and discrimination in a regulated industry. *American Economic Review*, *91*(4), 814–831.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, *13*(4), 412–438.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review*, *109*(10), 3395–3436.

Bolzendahl, C. I., & Myers, D. J. (2004). Feminist attitudes and support for gender equality: Opinion change in women and men, 1974–1998. *Social forces*, *83*(2), 759–789.

Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, *81*(2), 542–555.

Bornhorst, F., Ichino, A., Kirchkamp, O., Schlag, K. H., & Winter, E. (2010). Similarities and differences when building trust: the role of cultures. *Experimental Economics*, *13*(3), 260–283.

Boschini, A., Dreber, A., von Essen, E., Muren, A., Ranehill, E., et al. (2014). Gender and economic preferences in a large random sample. *Stockholm University, Department of Economics*.

Bottino, E., García-Muñoz, T., & Kujal, P. (2016). Gender biases in delegation. *ESI Working Paper 16-22*.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, *36*(1), 75–111.

Brandts, J., Groenert, V., & Rott, C. (2015). The impact of advice on women's and men's selection into competition. *Management Science*, *61*(5), 1018–1035.

Brewster, K. L., & Padavic, I. (2000). Change in gender-ideology, 1977–1996: The contributions of intracohort change and population turnover. *Journal of Marriage and Family*, *62*(2), 477–487.

Brock, J. M., & De Haas, R. (2019). Gender discrimination in small business landing. Evidence from a lab in-the-field experiment in Turkey. *EBRD Working Paper*.

Brooks, C., & Bolzendahl, C. (2004). The transformation of us gender role attitudes: Cohort replacement, social-structural change, and ideological learning. *Social Science Research*, *33*(1), 106–133.

Bryant, A. N. (2003). Changes in attitudes toward women's roles: Predicting gender-role traditionalism among college students. *Sex roles*, *48*(3), 131–142.

Burgess, S., & Ratto, M. (2003). The role of incentives in the public sector: Issues and evidence. *Oxford review of economic policy*, *19*(2), 285–300.

Burke, P. J. (1991). Identity processes and social stress. *American sociological review*, 836–849.

Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social forces*, *55*(4), 881–897.

Buser, T., Geijtenbeek, L., & Plug, E. (2018). Sexual orientation, competitiveness and income. *Journal of Economic Behavior & Organization*, *151*, 191–198.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447.

Buser, T., Niederle, M., & Oosterbeek, H. (2020). Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. *SSRN Electronic Journal*.

Buser, T., Niederle, M., & Oosterbeek, H. (2021). *Can competitiveness predict education and labor market outcomes? evidence from incentivized choice and survey measures* (Tech. Rep.). National Bureau of Economic Research.

Cadsby, C. B., Servátka, M., & Song, F. (2013). How competitive are female professionals? A tale of identity conflict. *Journal of Economic Behavior & Organization*, *92*, 284–303.

Cadsby, C. B., Song, F., & Tapon, F. (2007). Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal*, *50*(2), 387–405.

Cappelen, A. W., Falch, R., & Tungodden, B. (2020). Fair and unfair income inequality. *Handbook of Labor, Human Resources and Population Economics*, 1–25.

Cassar, A., Wordofa, F., & Zhang, Y. J. (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of*

Sciences, *113*(19), 5201–5205.

Castilla, C. (2018). Political role models and child marriage in india. *Review of Development Economics*, *22*(4), 1409–1431.

Castillo, M., & Petrie, R. (2010). Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior*, *68*(1), 50–59.

Chakraborty, P. (2019). It's a match! Or is it? A lab experiment on mentorship. *Mimeo*.

Charles, K. K., Guryan, J., & Pan, J. (2018). The effects of sexism on american women: The role of norms vs. discrimination. *National Bureau of Economic Research*.

Charness, G., Cobo-Reyes, R., Meraglia, S., & Sánchez, Á. (2020). Anticipated discrimination, choices, and performance: experimental evidence. *European Economic Review*, *127*, 103473.

Charness, G., Cobo-Reyes, R., Sanches, Á., & Meraglia, S. (2020). Anticipated discrimination, choices, and performance: Experimental evidence. *European Economic Review*.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica*, *72*(5), 1409–1443.

Ciabattari, T. (2001). Changes in men's conservative gender ideologies: Cohort and period influences. *Gender & Society*, *15*(4), 574–591.

Clain, S. H. (2000). Gender differences in full-time self-employment. *Journal of Economics and Business*, *52*(6), 499–513.

Cooper, D. J., & Weber, R. A. (2020). Recent advances in experimental coordination games. *Handbook of Experimental Game Theory*.

Corrigall, E. A., & Konrad, A. M. (2007). Gender role attitudes and careers: A longitudinal study. *Sex roles*, *56*(11), 847–855.

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies*, *75*(3), 729–762.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–74.

Cunningham, M., Beutel, A. M., Barber, J. S., & Thornton, A. (2005). Reciprocal relationships between attitudes about gender and social contexts during young adulthood. *Social Science Research*, *34*(4), 862–892.

Dal Bó, E., Finan, F., & Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, *128*(3), 1169–1218.

Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science*, *58*(11), 1982–2000.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, *51*(1), 816–835.

Delfgaauw, J., & Dur, R. (2007). Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, *62*(4), 605–624.

Deserranno, E. (2019). Financial incentives as signals: experimental evidence from the recruitment of village promoters in uganda. *American Economic Journal: Applied Economics*, *11*(1), 277–317.

Dhar, D., Jain, T., & Jayachandran, S. (2018). *Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india* (Tech. Rep.). National Bureau of Economic Research.

Dickinson, D., & Villeval, M.-C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic behavior*, *63*(1), 56–76.

Dimant, E. (2021). Hate trumps love: The impact of political polarization on social preferences.

Dohmen, T., & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, *101*(2), 556–90.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, *9*(3), 522–550.

Du, H., Xiao, Y., & Zhao, L. (2021). Education and gender role attitudes. *Journal of Population Economics*, *34*(2), 475–513.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–78.

Eastlund, T. (1998). Monetary blood donation incentives and the risk of transfusion-transmitted infection. *Transfusion*, *38*(9), 874–882.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games.

Economic Inquiry, *39*(2), 171–188.

Egan, M. L., Matvos, G., & Seru, A. (2017). When Harry fired Sally: The double standard in punishing misconduct. *National Bureau of Economic Research*.

Erat, S. (2013). Avoiding lying: The case of delegated deception. *Journal of Economic Behavior & Organization*, *93*, 273–278.

Eriksen, K. W., Kvaløy, O., & Luzuriaga, M. (2020). Risk-taking on behalf of others. *Journal of Behavioral and Experimental Finance*, *26*, 100283.

Erikson, E. H. (1994). *Identity and the life cycle*. WW Norton & Company.

Eriksson, T., & Villeval, M. C. (2008). Performance-pay, sorting and social motivation. *Journal of Economic Behavior & Organization*, *68*(2), 412–421.

European Commission. (2019). She figures 2018. *Directorate-General for Research and Innovation (European Commission)*.

Exley, C. L., Niederle, M., & Vesterlund, L. (2020). Knowing when to ask: The cost of leaning in. *Journal of Political Economy*, *128*(3), 816–854.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692.

Falk, A., Becker, A., Dohmen, T. J., Huffman, D., & Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Netspar discussion paper*.

Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, *362*(6412).

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, *96*(5), 1611–1630.

Fallucchi, F., Nosenzo, D., & Reuben, E. (2020). Measuring preferences for competition with experimentally-validated survey questions. *Journal of Economic Behavior & Organization*, *178*, 402–423.

Fallucchi, F., & Quercia, S. (2018). Affirmative action and retaliation in experimental contests. *Journal of Economic Behavior & Organization*, *156*, 23–40.

Fan, P.-L., & Marini, M. M. (2000). Influences on gender-role attitudes during the transition to adulthood. *Social Science Research*, *29*(2), 258–283.

Fehr, E., & List, J. A. (2004). The hidden costs and returns of incentives—trust and

trustworthiness among ceos. *Journal of the European Economic Association*, *2*(5), 743–771.

Fernández, R., Fogli, A., & Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, *119*(4), 1249–1299.

Fershtman, C., & Gneezy, U. (2001a). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, *116*(1), 351–377.

Fershtman, C., & Gneezy, U. (2001b). Strategic delegation: An experiment. *RAND Journal of Economics*, 352–368.

Fisk, S. R., & Overton, J. (2019). Who wants to lead? Anticipated gender discrimination reduces women's leadership ambitions. *Social Psychology Quarterly*, *82*(3), 319–332.

Fleming, P., & Zizzo, D. J. (2011). Social desirability, approval and public good contribution. *Personality and Individual Differences*, *51*(3), 258–262.

Flory, J. A., Leibbrandt, A., & List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, *82*(1), 122–155.

Friedl, A., Pondorfer, A., & Schmidt, U. (2020). Gender differences in social risk taking. *Journal of Economic Psychology*, *77*(102182).

Füllbrunn, S., Luhan, W. J., & Sanfeycd, A. (2020). Current issues in decision making for others. *Journal of Economic Psychology*, *77*(102250).

Gächter, S., & Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, *13*(3), 364–377.

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives*. Scribner.

Giamattei, M., Yahosseini, K. S., Gächter, S., & Molleman, L. (2020). Lioness lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, *6*(1), 95–111.

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, *127*(4), 1826–1863.

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, *77*(5), 1637–1664.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of economic perspectives*, *25*(4), 191–210.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791–810.

Görges, L. (2020). Of housewives and feminists: Gender norms and intra-household labour division. *Mimeo*.

Green, E., & Kloosterman, A. (2019). Agent sorting by incentive systems in mission firms: Implications for health care and other credence goods markets. *Available at SSRN 3497058*.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125.

Haaland, I., Roth, C., & Wohlfart, J. (2020). Designing Information Provision Experiments. *CAGE working paper no. 484*.

Hamman, J. R., Loewenstein, G., & Weber, R. A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship. *American Economic Review*, *100*(4), 1826–46.

Holm, H. J. (2000). Gender-Based Focal Points. *Games and Economic Behavior*, *32*(2), 292–314.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Ifcher, J., & Zarghamee, H. (2020). Do nominations close the gender gap in competition?

Inglehart, R. F., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., ... Puranen, B. (2014). Madrid, Spain Vienna, Austria: JD Systems Institute WVSA Secretariat.

Jacobsen, B., Lee, J. B., Marquering, W., & Zhang, C. Y. (2014). Gender differences in

optimism and asset allocation. *Journal of Economic Behavior & Organization*, *107*, 630–651.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, *7*(1), 63–88.

Jones, D., Tonin, M., & Vlassopoulos, M. (2018). Paying for what kind of performance? performance pay and multitasking in mission-oriented jobs. *CESifo Working Paper Series*.

Jost, J. T. (2019). A quarter century of system justification theory: Questions, answers, criticisms, and societal applications. *British Journal of Social Psychology*, *58*(2), 263–314.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27.

Kamas, L., & Preston, A. (2015). Can social preferences explain gender differences in economic behavior? *Journal of Economic Behavior & Organization*, *116*, 525–539.

Kane, E. W. (1995). Education and beliefs about gender inequality. *Social Problems*, *42*(1), 74–90.

Kessel, D., Mollerstrom, J., & van Veldhuizen, R. (2019). Can simple advice eliminate the gender gap in willingness to compete? *Working Paper; Available at https://www.aeaweb.org/conference/2020/preliminary/paper/GbaKAeNY (Accessed 12.4.2021)*.

Kranton, R. E. (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review*, *106*(5), 405–09.

Lagarde, M., & Blaauw, D. (2014). Pro-social preferences and self-selection into jobs: Evidence from south african nurses. *Journal of Economic Behavior & Organization*, *107*, 136–152.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361.

Leibbrandt, A., Wang, L. C., & Foo, C. (2018). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, *64*(8), 3501–3516.

Mason, M. A., Wolfinger, N. H., & Goulden, M. (2013). *Do babies matter? gender and family in the ivory tower*. Rutgers University Press.

Mellström, C., & Johannesson, M. (2008). Crowding out in blood donation: was titmuss right? *Journal of the European Economic Association*, *6*(4), 845–863.

Moore, L. M., & Vanneman, R. (2003). Context matters: Effects of the proportion of fundamentalists on gender attitudes. *Social Forces*, *82*(1), 115–139.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, *119*(1), 39–77.

Niederle, M. (2015). *Gender. Handbook of Experimental Economics.* Princeton University Press.

Niederle, M. (2017). A gender agenda: A progress report on competitiveness. *American Economic Review*, *107*(5), 115–119.

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1–16.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica*, *70*(3), 971–1005.

OECD. (2019). *Pisa 2018 results (volume i): What students know and can do*. OECD.

Office of the Registrar General & Census Commissioner. (2011). Census Data. https://censusindia.gov.in/2011-common/censusdata2011.htm.

Postlewaite, A. (2011). Social norms and preferences, chapter for the handbook for social economics edited by j. benhabib, a. bisin and m. jackson. In *Handbook of social economics* (Vol. 1, pp. 31–67). Elsevier.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403–4408.

Reuben, E., Sapienza, P., & Zingales, L. (2015). Competitiveness and the gender gap among young business professionals. *CEPR Discussion Paper No. DP10924*.

Reuben, E., Wiswall, M., & Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, *127*(604), 2153–2186.

Saccardo, S., Pietrasz, A., & Gneezy, U. (2018). On the size of the gender difference in competitiveness. *Management Science*, *64*(4), 1541–1554.

Sardinha, L., & Catalán, H. E. N. (2018). Attitudes towards domestic violence in 49 low-

and middle-income countries: A gendered analysis of prevalence and country-level correlates. *PloS one*, *13*(10).

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review*, *107*(5), 141–145.

Serra, D., Serneels, P., & Barr, A. (2011). Intrinsic motivations and the non-profit health sector: Evidence from ethiopia. *Personality and Individual Differences*, *51*(3), 309–314.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, *97*(3), 999–1012.

Slovic, P., Weinstein, M. S., & Lichtenstein, S. (1967). Sex differences in the risks a person selects for himself and the risks he selects for someone else. *Research Bulletin*, *7*(10).

Solnick, S. J. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*(2), 189–200.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.

Stöber, J. (2001). The social desirability scale-17 (sds-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, *17*(3), 222.

Stryker, S. (1968). Identity salience and role performance: The relevance of symbolic interaction theory for family research. *Journal of Marriage and the Family*, 558–564.

Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining—beware the same sex! *Experimental Economics*, *12*(3), 318–331.

Sutter, M., & Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, *61*(10), 2339–2354.

Tallichet, S. E., & Willits, F. K. (1986). Gender-role attitude change of young women: Influential factors from a panel study. *Social Psychology Quarterly*, 219–227.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American economic review*, *93*(2), 175–179.

Thornton, A., Alwin, D. F., & Camburn, D. (1983). Causes and consequences of sex-role

attitudes and attitude change. *American Sociological Review*, 211–227.

Tungodden, J. (2019). When parents decide: Gender differences in competitiveness. *Mimeo*.

van Veldhuizen, R. (2021). Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness? *Mimeo*.

World Economic Forum. (2020). Global gender gap report. http://www3.weforum.org/docs/WEF_GGGR_2020.pdf.

World data atlas. (2020). https://knoema.com/ATLAS.

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2016). The effect of feedback on gender differences in competitive choices. *Available at SSRN 1976073*.

Instructions

Chapter II - Stage 1

General

Please take your time to read the following instructions carefully.

This study consists of several questions and should take approximately 10 minutes to complete.

At the end of the study, we will randomly select 5 participants and pay them £50 each.

Only participants who complete the study will be eligible for payment.

Please do not close this window or use the “back” button of your browser at any point during the study.

If you do so, you will exit the study and you will not be able to resume participation.

Preferences

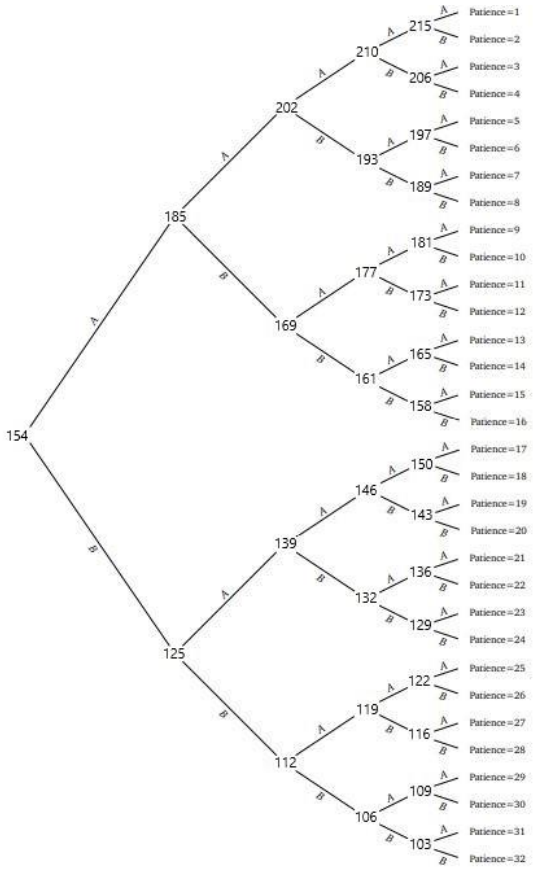
In the following, “willingness to act” indicates the following introduction: We now ask for your willingness to act in a certain way in four different areas. Please again indicate your answer on a scale from 0 to 10, where 0 means you are “completely unwilling to do so” and a 10 means you are “very willing to do so”. You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Similarly, “self-assessments” indicate that the respective statement was preceded by the following introduction: How well do the following statements describe you as a person? Please indicate your answer on a scale from 0 to 10. A 0 means “does not describe me at all” and a 10 means “describes me perfectly”. You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Patience

1. (Sequence of five interdependent quantitative questions:) Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e, future prices arethe same as today’s prices. Please consider the following: Would you rather receive 100 pounds today or x pounds in 12 months?

The precise sequence of questions was given by the “tree” logic in the following figure.



Tree for the staircase time task (numbers = payment in 12 months, A = choice of "100 pounds today", B = choice of "x pounds in 12 months". The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 100 pounds today or 154 pounds in 12 months from now (leftmost decision node). In case the respondent opted for the payment today ("A"), in the second question the payment in 12 months was adjusted upwards to 185 pounds. If, on the other hand, the respondent chose the payment in 12 months, the corresponding payment was adjusted down to 125 pounds. Working further through the tree follows the same logic.

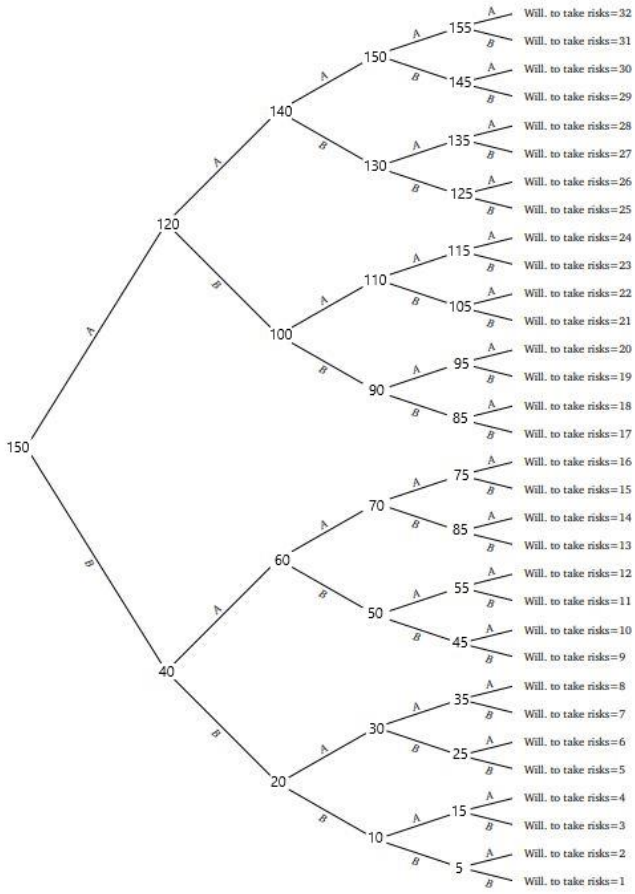
2. (Willingness to act:) How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

Risk taking

1. (Similar to self-assessment:) Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

2. (Sequence of five interdependent quantitative questions:) Please imagine the following situation. You can choose between a sure payment of a particular amount of money, or a draw, where you would have an equal chance of getting amount x or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving amount x, and the same 50 percent chance of receiving nothing, or the amount of y as a sure payment?

   The precise sequence of questions was given by the "tree" logic in following figure.



Tree for the staircase risk task (numbers = sure payment, A = choice of lottery, B = choice of sure payment). The staircase procedure worked as follows. First, each respondent was asked whether they would prefer to receive 150 pounds for sure or whether they preferred a 50:50 chance of receiving 300 pounds or nothing. In case the respondent opted for the safe choice ("B"), the safe amount of money being offered in the second question decreased to 80 pounds. If, on the other hand, the respondent opted for the gamble ("A"), the safe amount was increased to 240 pounds. Working further through the tree follows the same logic.

Positive reciprocity

1. (Self-assessment:) When someone does me a favor I am willing to return it.

2. (Hypothetical situation:) Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about 12 pounds in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs 5 pounds, the most expensive one costs 30 pounds. Do you give one of the presents to the stranger as a “thank-you”- gift? If so, which present do you give to the stranger? No present / The present worth 3 / 6 / 9 / 12 / 15 / 18 pounds.

Negative reciprocity

1. (Self-assessment:) If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.

2. (Willingness to act:) How willing are you to punish someone who treats you unfairly, even if there may be costs for you?

3. (Willingness to act:) How willing are you to punish someone who treats others unfairly, even if there may be costs for you?

Altruism

1. (Hypothetical situation:) Imagine the following situation: Today you unexpectedly received 700 pounds. How much of this amount would you donate to a good cause? (Values between 0 and 700 are allowed.)

2. (Willingness to act:) How willing are you to give to good causes without expecting anything in return?

Trust

1. (Self-assessment:) I assume that people have only the best intentions.

Social desirability

Please read each statement carefully and decide if that statement describes you or not. If it describes you, click on the word "true"; if not, click on the word "false".

1. I sometimes drop litter rather than carry it to a bin.

2. I always admit my mistakes openly and face potential negative consequences.
3. In traffic I am always polite and considerate of others.
4. I always accept others' opinions, even when they don't agree with my own.
5. I take out my bad moods on others now and then.
6. There has been an occasion when I took advantage of someone else.
7. In conversations I always listen attentively and let others finish their sentences.
8. I never hesitate to help someone in case of emergency.
9. When I have made a promise, I keep it – no ifs, ands or buts.
10. I occasionally speak badly of others behind their back.
11. I would never live off other people.
12. I always stay friendly and courteous with other people, even when I am stressed out.
13. During arguments I always stay objective and matter-of-fact.
14. There has been at least one occasion when I failed to return an item that I borrowed.
15. I always eat a healthy diet.
16. Sometimes I only help because I expect something in return.

Self-control

Please state the extent to which you agree with the following statements.

Please use the scale below which goes from 1 to 5, where 1 means you "strongly disagree" and 5 means you "strongly agree".

You can also use any numbers between 1 and 5 to indicate where you fall on the scale, like 1, 2, 3, 4, 5.

1. I am good at resisting temptation.
2. I have a hard time breaking bad habits.
3. I am lazy.

4. I say inappropriate things.
5. I do certain things that are bad for me, if they are fun.
6. I refuse things that are bad for me.
7. I wish I had more self-discipline.
8. People would say that I have an iron self-discipline.
9. Pleasure and fun sometimes keep me from getting work done.
10. Pleasure and fun sometimes keep me from getting work done.
11. I am able to work effectively toward long-term goals.
12. Sometimes I can't stop myself from doing something, even if I know it is wrong.
13. I often act without thinking through all the alternatives.

Chapter II - Stage 2 (General Introduction)

Introduction

Thank you for taking part in this study.

You will shortly receive detailed instructions for this study directly on your computer screen. The instructions will explain how you can earn money in today's experiment. It is therefore important that you read the instructions carefully.

If you have any questions please let us know by raising your hand. Your question will then be answered at your cubicle.

In the experiment today you will asked to sum a series of numbers. You must make the sums **without any helping device, i.e., pocket calculators or cell phones, are not allowed**. If you use any helping device, you will be excluded from the experiment and from all payments. However, you are allowed to use the pen and paper positioned next to your keyboard.

Please put away your phones now.

Finally, please note that you are not allowed to communicate with other participants throughout the study and that unnecessary interference with computers will lead to exclusion from the experiment.

Are there any questions at this stage?

If not, we are ready to start.

You can now press the Continue button on your screen.

Chapter II - Stage 2

General

Thank you for taking part in this study.

Please read the following information carefully.

If you have any questions please let us know by raising your hand. Your question will then be answered at your cubicle.

During the study, you can earn money by solving tasks. You will receive £4.00 for your participation.

The additional amount of money that you will earn depends on your decisions. At the end of this study, you will receive the money that you have earned in cash.

The study consists of two parts. You will be paid only for one part, either Part 1 or Part 2, selected at random. Each part has an equal chance of being selected.

Please do not close the last page of the study as it will determine your payment.

Each part will be introduced on a screen with the header "Instructions". These instructions will explain in detail what the respective part of the study is about.

The tasks will involve summing numbers. You must solve the problems without any helping device, i.e., devices such as paper and pencil, pocket calculators or cell phones, are not allowed. If you use any helping device, you will be immediately excluded from the experiment and from all payments.

Please note that communication between participants is strictly prohibited. In addition, we would like to point out that you may only use the computer functions which are required for the experiment.

Communication between participants will lead to exclusion from the experiment.

Instructions - Part 1

You have 5 minutes to complete the following task.

Your task is to find the highest number contained in each of two matrices, add them up and report the sum in a box.

If you reported the correct answer, once you click the "Submit" button the next pair of matrices will appear.

Otherwise, you will be shown the same pair of matrices until you provide the correct answer.

You will earn £0.20 for each pair of matrices you solve correctly.

On the next page, you will find an example.

You can see an example of the matrices you will be asked to solve below.

Your task is to find the highest number in the left matrix, that is 76.

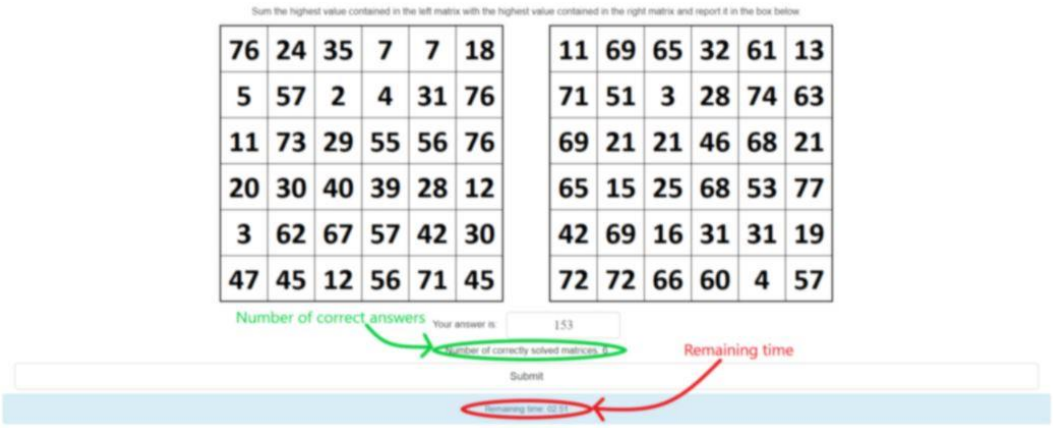
Then find the highest number in the right matrix, that is 77.

Sum them, 76+77= 153.

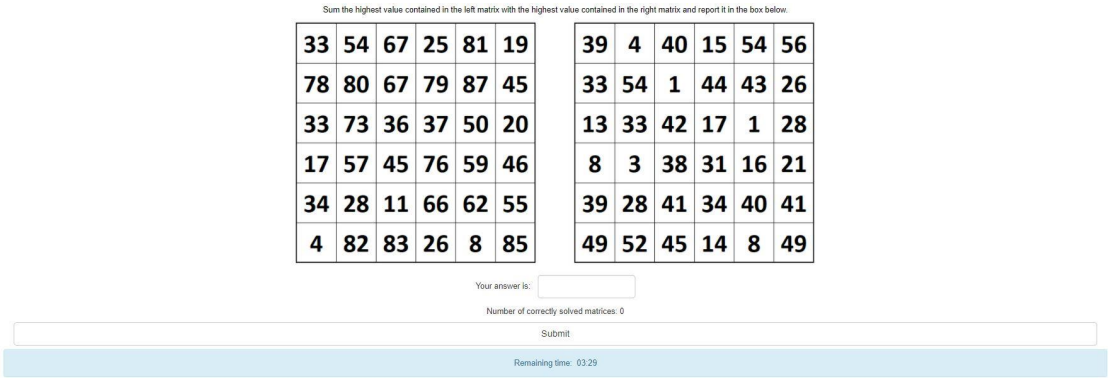
Report the value in the input box and click "Submit".

If you provided the correct answer the next matrices will appear until the time is over.

On your screen you will also see the number of correctly solved matrices so far and the remaining time to complete the task.



The task



Follow-up questions

1. On a scale from 1 (none at all) to 5 (very much), how much effort did you exert solving the matrices?

2. On a scale from 1 (not at all) to 5 (very much), how stressed did you feel?

3. On a scale from 1 (not at all) to 5 (very much), how exhausted did you get?

Instructions - Part 2

In the second part of the study, your task is again to solve matrices.

You will have 10 minutes to complete the task.

This time, however, you can choose whether to complete the task for your own benefit or for the benefit of a charity (you can choose one of the following charities: Amnesty International, Greenpeace, The National Trust, Oxfam, Red Cross, Save the Children, Cancer Research UK).

This choice determines who receives the money you generate in the task.

At the end of the task, we will count the number of matrices that you solved correctly. For each correctly solved pair of matrices, you will generate £0.05.

If you choose to work for yourself, you will keep this amount for yourself.

If you choose to work for the charity, we will donate twice this amount on your behalf to a charity of your choice.

Treatment 1

Moreover, your choice whether to complete the task for your own benefit or for the benefit of the charity determines your payment from completing the task.

If you choose to work for yourself, you will be paid £5.00 if you solve at least 17 matrices, and £0.00 otherwise.

If you choose to work for the charity, you will be paid £3.00 regardless of how many matrices you solve.

Which option do you choose?

If you choose Work for Myself, you will earn:

£0.05 for each matrix you solve correctly, which you keep for yourself.

Plus a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

If you choose Work for a Charity, you will earn:

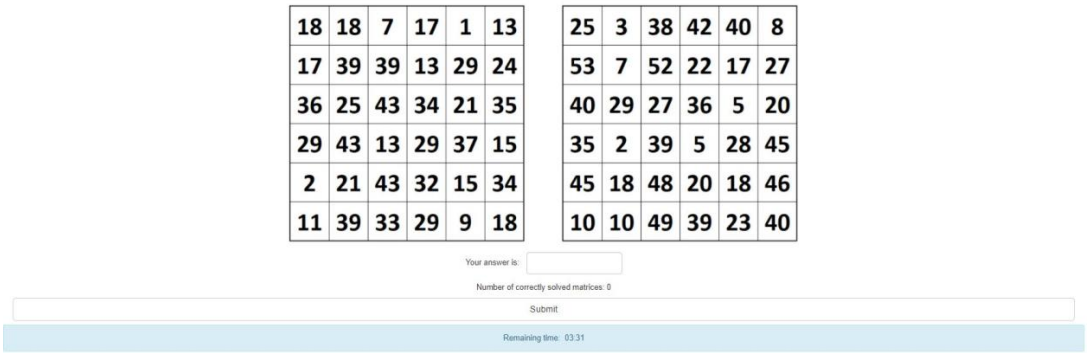
£0.05 for each matrix you solve correctly, which are multiplied by 2 and donated to a charity of your choice.

Plus £3.00 for yourself independently of the number of matrices you manage to solve correctly.

The task

Sum the highest value contained in the left matrix with the highest value contained in the right matrix and report it in the box below.

You are now working for Yourself/a Charity.



Treatment 2

Moreover, your choice whether to complete the task for your own benefit or for the benefit of the charity determines your payment from completing the task.

If you choose to work for yourself, you will be paid £5.00 if you solve at least 17 matrices, and £0.00 otherwise.

If you choose to work for the charity, you will have an additional choice to make:

1) you can choose to be paid £3.00 regardless of how many matrices you solve;

or

2) you can choose to be paid £5.00 if you solve at least 17 matrices, and £0.00 otherwise.

Which option do you choose?

If you choose Work for Myself, you will earn:

£0.05 for each matrix you solve correctly, which you keep for yourself.

Plus a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

If you choose Work for a Charity, you will earn:

£0.05 for each matrix you solve correctly, which are multiplied by 2 and donated to a charity of your choice.

Plus either £3.00 for yourself independently of the number of matrices you manage to solve correctly; or a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

(You will be able to choose between these options on the next screen)

You chose Work for a Charity.

Which payment option do you choose?

£0.05 for each matrix you solve correctly, which are multiplied by 2 and donated to a charity of your choice.

Plus £3.00 for yourself independently of the number of matrices you manage to solve correctly.

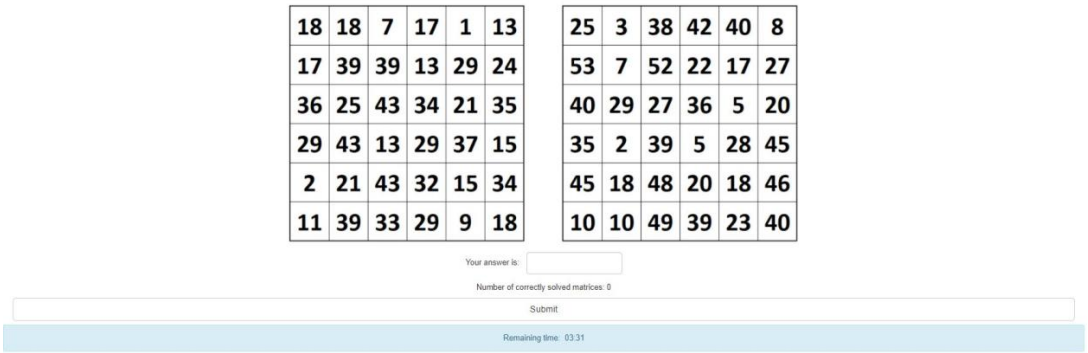
£0.05 for each matrix you solve correctly, which are multiplied by 2 and donated to a charity of your choice.

Plus a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

The task

Sum the highest value contained in the left matrix with the highest value contained in the right matrix and report it in the box below.

You are now working for Yourself/a Charity.



Treatment 3

Moreover, regardless of whether you choose to complete the task for your own benefit or for the benefit of the charity, you will be paid £5.00 if you solve at least 17 matrices, and £0.00 otherwise.

Which option do you choose?

If you choose Work for Myself, you will earn:

£0.05 for each matrix you solve correctly, which you keep for yourself.

Plus a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

If you choose Work for a Charity, you will earn:

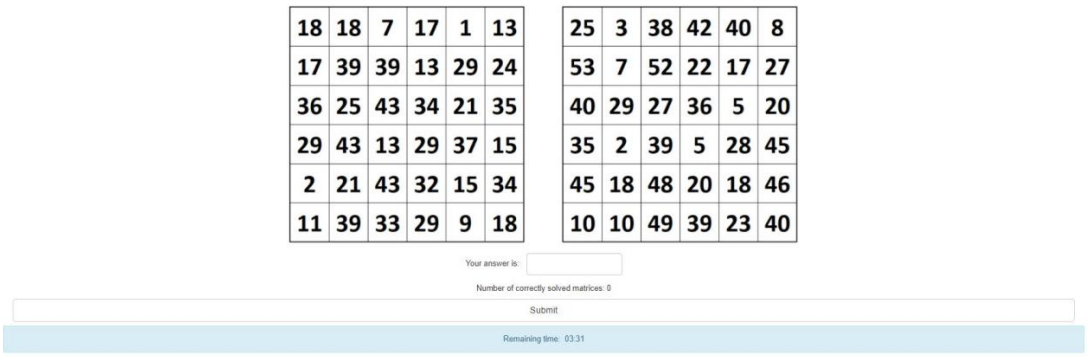
£0.05 for each matrix you solve correctly, which are multiplied by 2 and donated to a charity of your choice.

Plus a bonus of £5.00 if you manage to solve correctly at least 17 matrices, and £0.00 otherwise.

The task

Sum the highest value contained in the left matrix with the highest value contained in the right matrix and report it in the box below.

You are now working for Yourself/a Charity.



Follow-up questions

1. On a scale from 1 (none at all) to 5 (very much), how much effort did you exert solving the matrices?

2. On a scale from 1 (not at all) to 5 (very much), how stressed did you feel?

3. On a scale from 1 (not at all) to 5 (very much), how exhausted did you get?

Chapter III

OWNDECISION

Introduction

Thank you for taking part in this study.

Please read the following information carefully.

If you have any questions, please let us know by raising your hand. Your question will then be answered in your cubicle.

You will receive **£4.00** for your participation.

The additional amount of money that you will **earn depends on your decisions and that of other players in the study.**

At the end of this study, you will receive the money that you have earned in cash.

The study consists of four parts. You will be paid only for one part, which will be selected at random at the end of the study. Each part has an equal chance of being selected.

Please do not close the last page of the study as it will determine your payment.

Each task will be introduced on your screen with the header **"Instructions".** The instructions will explain in detail what the respective part of the study is about.

One of the tasks involves summing numbers. You must solve the problems **without any helping device, i.e., devices such as pocket calculators or cell phones, are not allowed.** If you use any helping device, you will be immediately excluded from the experiment and from all payments. However, you are **allowed to use the pen and paper positioned next to your keyboard.**

Finally, please note that you are not allowed to communicate with other participants throughout the study and that unnecessary interference with computers will lead to exclusion from the experiment.

Question 1

What is your sex? [Options: Male, Female]

Question 2

How old are you? [Options: numeric input _____]

Question 3

What is your country of residence? [Options: United Kingdom, Other]

i) **Stage 1**

Instructions - Task 1

You have **5 minutes** to complete the following task.

For Task 1, you will be asked to **calculate the sum of five randomly chosen numbers**.

Once you **click the Submit button**, the next numbers will appear.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

Your answers to the problems are anonymous.

If Task 1 is the one randomly selected for payment, then you **get £0.50 per problem you solve correctly** in the **5 minutes**.

Your payment does not decrease if you provide an incorrect answer to a problem.

We refer to this payment as the **piece rate payment**.

On the next page, you will find **an example.**

Below you can see an **example** of the sums you will be asked to make.

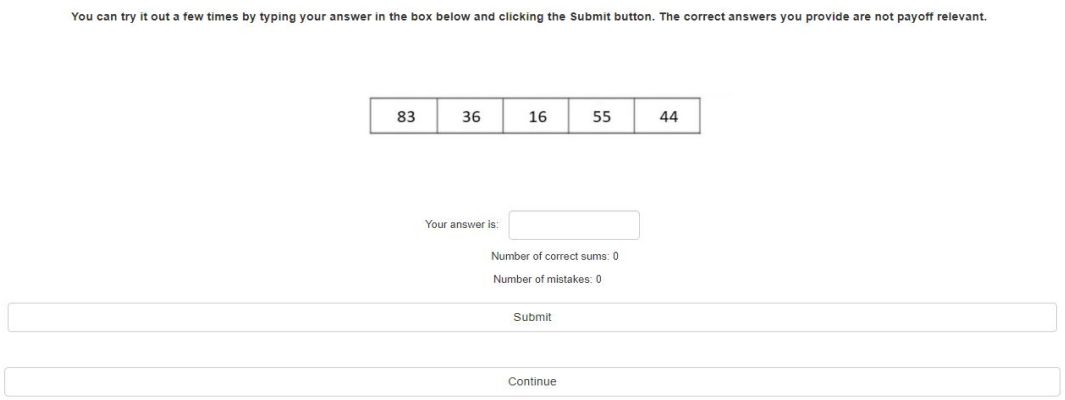
Your task is to add the numbers contained in the image:

83+36+16+55+44 = 234

You submit an answer by clicking the Submit button with your mouse.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

On your screen, you will also see the number of correctly solved sums and the remaining time.



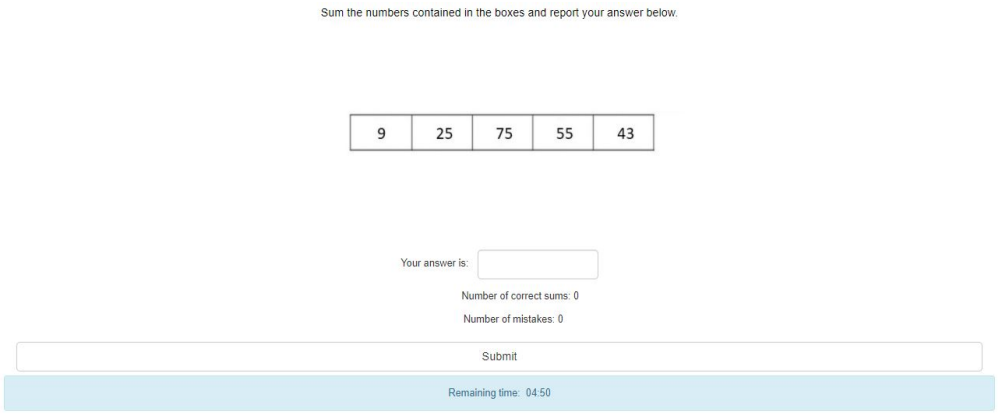
The example is over.

Please press the Continue button below.

Click the Start button if you are ready to start Task 1.

Remember, you have **5 minutes** to solve this task.

Task 1



Bonus payment

You have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if your answer is correct.

How well did you perform in Task 1 relative to the **other subjects in this room**?

[Options:1 = best 25%, 2, 3, 4 = worst 25%]

ii) Stage 2

Instructions - Task 2

Again, you have **5 minutes** to complete the following task.

As in Task 1, you will be given **5 minutes** to calculate the correct sum of a series of five numbers. However, for this task, your payment depends on your performance relative to that of a group of other participants.

Each group consists of four people (2 males and 2 females), the three other members of your group are located in the same row as you. That is, the people immediately next to you are part of your group.

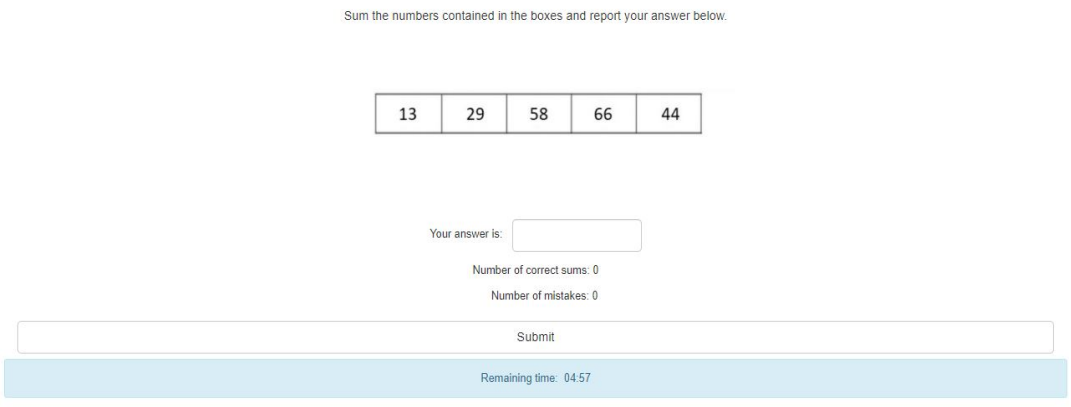
If Task 2 is the one randomly selected for payment, then **your earnings depend on the number of problems you solve compared to the other people in your group.** The individual who correctly solves the largest number of problems will receive £2 per correct answer, while the other participants receive **no payment** for this task. If there are ties, the winner will be randomly determined.

We refer to this as **the tournament payment.**

Click the Start button if you are ready to start Task 2.

Remember, you have **5 minutes** to solve this task.

Task 2



Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

How well did you perform in Task 2 relative to the **other subjects in your group**?

[Options:1 = best, 2, 3, 4 = worst]

iii) Stage 3

Instructions - Task 3

Again, you have **5 minutes** to complete the following task.

As in Task 1 and Task 2, you will be given **5 minutes** to calculate the correct sum of a series of five 2 numbers. However, you will now **get to choose which of the two previous payment schemes, piece rate** or **tournament**, **you prefer to apply to your performance on the third task.**

If Task 3 is the one randomly selected for payment, then your earnings for this task are determined as follows.

- If you choose the **piece-rate payment**, you receive **£0.50 per problem you solve correctly.**

Your payment does not decrease if you provide an incorrect answer to a problem.

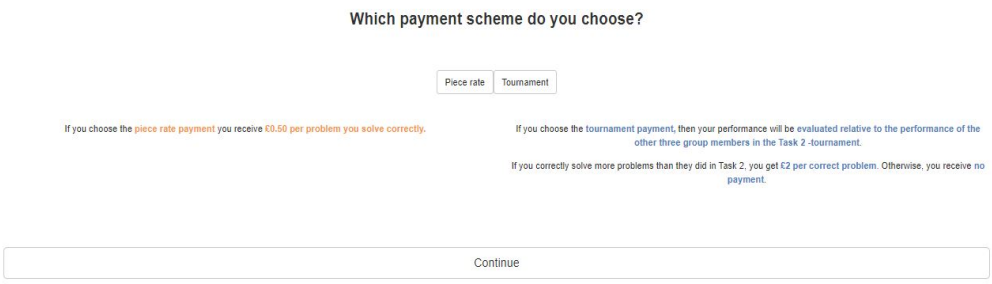
- If you choose the **tournament payment**, then your performance in Task 3 will be **evaluated relative to the performance of the other three group members in the Task 2 -tournament** (the Task 2-tournament is the one you just completed).

If you correctly solve more problems than your group members did in Task 2, then you receive four times the payment from the piece rate, which is **£2 per correct problem**. However, you will receive **no earnings** for this task if you choose the tournament and you do not solve more problems correctly in Task 3 than the other group members did in the Task-2 tournament. You will not be informed of how you did in the tournament until all four tasks have been completed

If there are ties, the winner will be randomly determined.

Your group members are the same as in the previous stage.

Task 3 – Choice screen



[Depending on what subjects choose, they will either see the next or the second next screen.]

You chose: **Piece rate payment**

You will get £0.50 per problem you solve correctly.

Your payment does not decrease if you provide an incorrect answer to a problem.

You chose: **Tournament payment**

Your performance in Task 3 will be **evaluated relative to the performance of the other three group members in the Task 2 -tournament** (the Task 2-tournament is the one you just completed).

If you correctly solve more problems than your group members did in Task 2, then you receive four times the payment from the piece rate, which is **£2 per correct problem.** However, you will receive **no earnings** for this task if you choose the tournament and you do not solve more problems correctly in Task 3 than the other group members did in the Task-2 tournament. You will not be informed of how you did in the tournament until all four tasks have been completed

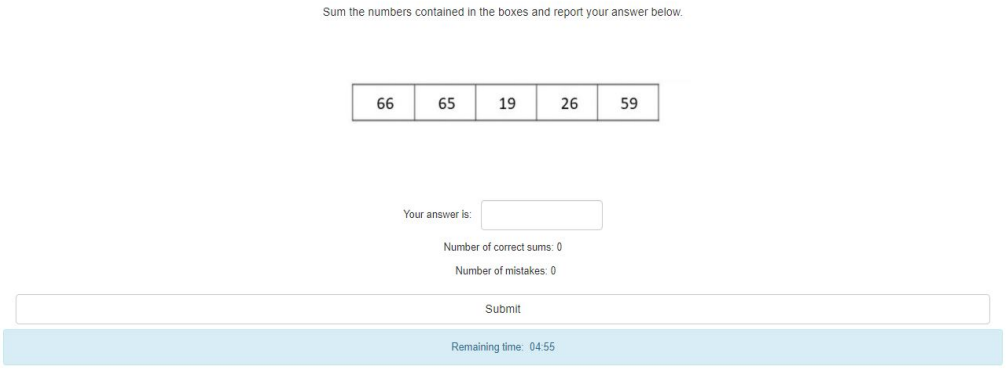
If there are ties, the winner will be randomly determined.

Your group members are the same as in the previous stage.

Click the Start button if you are ready to start Task 3.

Remember, you have **5 minutes** to solve this task.

Task 3



iv) Stage 4

Instructions - Task 4

For Task 4, you will be asked to make a series of **ten pairwise decisions** between **"Option A"** and **"Option B".**

If Task 4 is the one selected for payment, **your earnings depend on which of the 10 decisions is randomly selected**, you will be paid out according to this decision.

On the next page, you will find an example.

This is how the decision sheet will look like:

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

Now, please look at the **first decision:**

Option A has a 1 in 10 chance of paying **£2.20** and a 9 in 10 chance of paying **£1.75**

Option B has a 1 in 10 chance of paying **£4.20** and a 9 in 10 chance of paying **£0.10**

The other decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for the **last decision** in the bottom row, a random draw will not be needed since each option pays the highest payoff for sure, so your choice here is between £2.20 and £4.20.

To summarise, you will make ten choices: for each decision row, you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows.

Please click Continue if you have understood the instructions and wish to continue to the Decision table.

Click the Start button if you are ready to start Task 4.

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

v) Final questionnaire part 1

Please answer the following questions

Page 1/2

Question 1

On a scale from **0 (not at all like me)** to **10 (exactly like me),** please indicate the extent to which the following statement **describes you**.

Competition brings the best out of me.

[Options: 0 (not at all like me), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (exactly like me)]

Question 2

On a scale from **0 (completely unwilling to take risks)** to **10 (very willing to take risks)**, in general, how willing are you to **take risks**?

[Options: 0 (completely unwilling to take risks),1,2,3,4,5,6,7,8,9,10 (very willing to take risks)]

Question 3

On a scale from **1 (very poor)** to **7 (excellent),** how would you classify your maths skills?

[Options: 1(very poor),1,2,3,4,5,6, 7 (excellent)]

vi) Final questionnaire part 2

Please answer the following questions

Page 2/2

Question 1

How religious do you consider yourself to be?

[Options: Very religious, Moderately religious, Not religious]

Question 2

What is your political identity?
[Options: Strongly conservative, Somewhat conservative, Neutral, Somewhat liberal, Strongly liberal]

Question 3

What's your country/region of primary citizenship?

[Options: text input _____]

Question 4

What is your field of study?

[Options: text input _____]

Earnings

Please don't close this window.

Payment for **Task 1**: [Stage 1]

Payment for **Task 2**: [Stage 2]

Payment for **Task 3**: [Stage 3]

Payment for **Task 3**: [Stage 4]

The task chosen for payment is number: [Randomly chosen task]

This means that your total earnings are:

[£4 (participation fee) + Randomly chosen task + random bonus]

Please remain seated. The experimenter will soon come and pay you at your cubicle.

Thank you for taking part in this study.

GENDERINFO & NOGENDERINFO

Agents

[Please note that the instructions for the agents are the same in *GenderInfo* and in *NoGenderInfo*.]

Introduction

Thank you for taking part in this study.

Please read the following information carefully.

If you have any questions, please let us know by raising your hand. Your question will then be answered in your cubicle.

You will receive **£4.00** for your participation.

The additional amount of money that you will **earn depends on your decisions and that of other players in the study.**

At the end of this study, you will receive the money that you have earned in cash.

The study consists of four parts. You will be paid only for one part, which will be selected at random at the end of the study. Each part has an equal chance of being selected.

Please do not close the last page of the study as it will determine your payment.

Each task will be introduced on your screen with the header **"Instructions".** The instructions will explain in detail what the respective part of the study is about.

One of the tasks involves summing numbers. You must solve the problems **without any helping device, i.e., devices such as pocket calculators or cell phones, are not allowed.** If you use any helping device, you will be immediately excluded from the experiment and from all payments. However, you are **allowed to use the pen and paper positioned next to your keyboard.**

Finally, please note that you are not allowed to communicate with other participants throughout the study and that unnecessary interference with computers will lead to exclusion from the experiment.

Question 1

What is your sex? [Options: Male, Female]

Question 2

How old are you? [Options: numeric input: _____]

Question 3

What is your country of residence? [Options: United Kingdom, Other]

i) **Stage 1**

Instructions - Task 1

You have **5 minutes** to complete the following task.

For Task 1, you will be asked to **calculate the sum of five randomly chosen numbers**.

Once you **click the Submit button**, the next numbers will appear.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

Your answers to the problems are anonymous.

If Task 1 is the one randomly selected for payment, then you **get £0.50 per problem you solve correctly** in the **5 minutes**.

Your payment does not decrease if you provide an incorrect answer to a problem.

We refer to this payment as the **piece rate payment**.

On the next page, you will find **an example.**

Below you can see an **example** of the sums you will be asked to make.

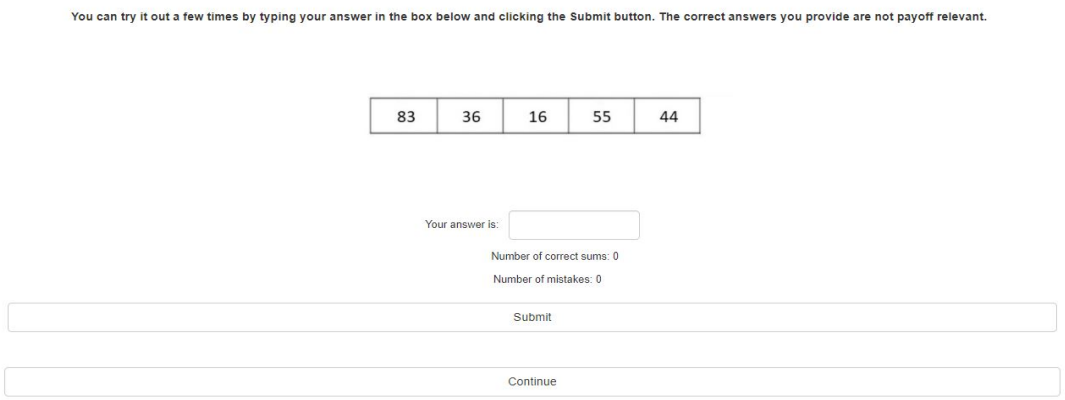
Your task is to add the numbers contained in the image:

83+36+16+55+44 = 234

You submit an answer by clicking the Submit button with your mouse.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

On your screen, you will also see the number of correctly solved sums and the remaining time.



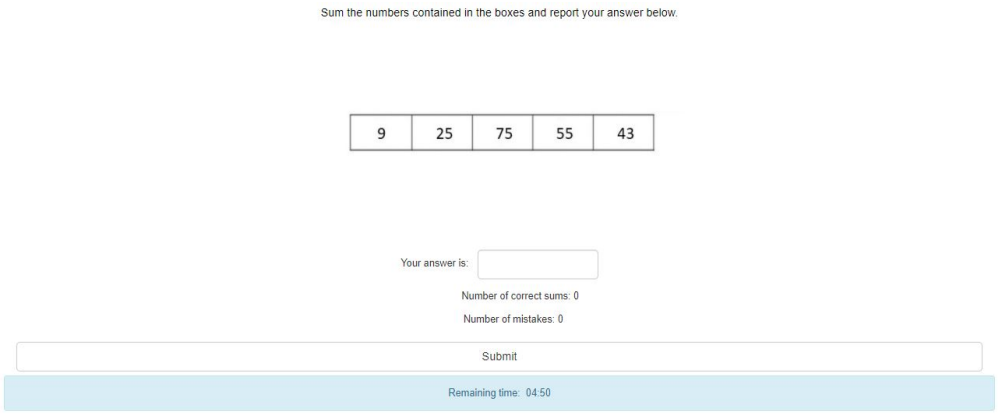
The example is over.

Please press the Continue button below.

Click the Start button if you are ready to start Task 1.

Remember, you have **5 minutes** to solve this task.

Task 1



Bonus payment

You have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if your answer is correct.

How well did you perform in Task 1 relative to the **other subjects in this room**?

[Options:1 = best 25%, 2, 3, 4 = worst 25%]

ii) **Stage 2**

Instructions - Task 2

Again, you have **5 minutes** to complete the following task.

As in Task 1, you will be given **5 minutes** to calculate the correct sum of a series of five numbers. However, for this task, your payment depends on your performance relative to that of a group of other participants.

Each group consists of four people (2 males and 2 females), the three other members of your group are located in the same row as you. That is, the people immediately next to you are part of your group.

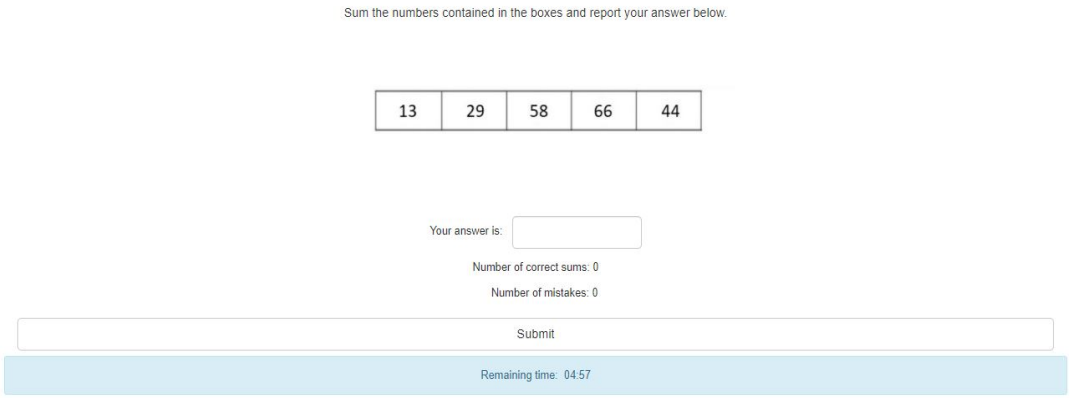
If Task 2 is the one randomly selected for payment, then **your earnings depend on the number of problems you solve compared to the other people in your group.** The individual who correctly solves the largest number of problems will receive £2 per correct answer, while the other participants receive **no payment** for this task. If there are ties, the winner will be randomly determined.

We refer to this as **the tournament payment.**

Click the Start button if you are ready to start Task 2.

Remember, you have **5 minutes** to solve this task.

Task 2



Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

How well did you perform in Task 2 relative to the **other subjects in your group**?

[Options:1 = best, 2, 3, 4 = worst]

i) **Stage 3**

Instructions - Task 3

Instructions - Task 3

Again, you have **5 minutes** to complete the following task.

For this task, you and the other participants have been randomly assigned one of two roles, Player A or Player B.

You are Player A.

As in Task 1, you will be given **5 minutes** to calculate the correct sum of a series of five numbers. However, this time **your payment scheme depends on the choice of another randomly matched participant assigned the role of Player B.** He/she will choose which of the two payment schemes, **piece rate** or **tournament**, he/she wants to apply to your performance in Task 3.

Your identity is kept anonymous.

If Task 3 is randomly selected for payment, **your earnings depend on which type of payment is chosen for you.**

Instructions continue on the next page.

Instructions - Task 3

- If Player B chooses the **piece-rate payment**, then you get **£0.50 per problem you solve correctly** in the 5 minutes.

Your payment does not decrease if you provide an incorrect answer to a problem.

- If Player B chooses the **tournament payment**, then **your performance in Task 3 will be evaluated relative to the performance of the other three group members** in the Task 2 - tournament (the Task 2-tournament is the one you just completed).

If you correctly solve more problems than your group members did in Task 2, then you receive four times the payment from the piece rate, which is **£2 per correct problem**. However, you will receive **no earnings** for this task if Player B chooses the tournament and you do not solve more problems correctly in Task 3 than the other group members did in the Task-2 tournament.

If there are ties, the winner will be randomly determined.

Your group members are the same as in the previous task.

Please answer the following questions

Question 1

If given the choice, which payment scheme would you chose for yourself?

[Options: Piece-rate, Tournament]

Question 2

Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

Which payment scheme will **Player B** choose for you?

[Options: Piece-rate, Tournament]

On the next page, you will find out under which payment scheme you will be working in Task 3.

[Depending on what the principals chose, the agents will either see the next or the second next screen.]

Player B chose: **Piece rate payment**

You will get **£0.50 per problem you solve correctly.**

Your payment does not decrease if you provide an incorrect answer to a problem.

Player B chose: **Tournament payment**

Your performance in Task 3 will be **evaluated relative to the performance of the other three group members in the Task 2 -tournament** (the Task 2-tournament is the one you just completed).

If you correctly solve more problems than your group members did in Task 2, then you receive four times the payment from the piece rate, which is **£2 per correct problem**. However, you will receive **no earnings** for this task if he/she chooses the tournament and you do not solve more problems correctly in Task 3 than the others in your group did in the Task-2 tournament. You will not be informed of how you did in the tournament until all four tasks have been completed.

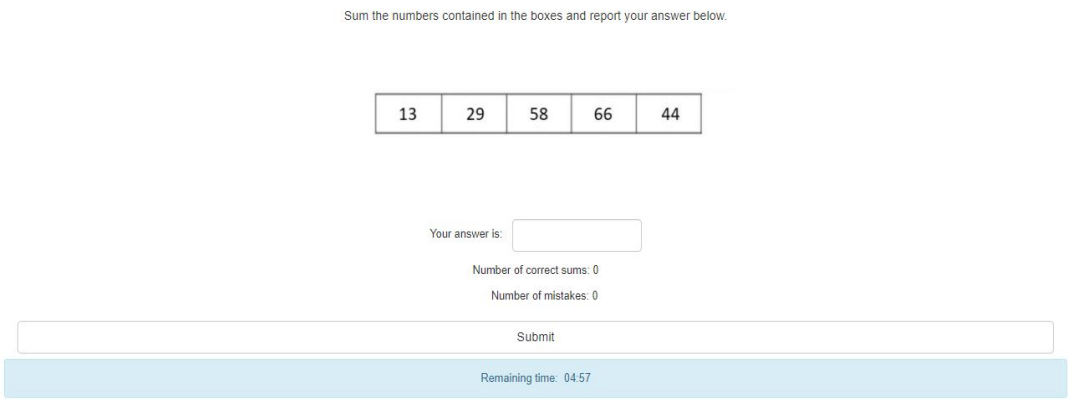
If there are ties, the winner will be randomly determined.

Your group members are the same as in the previous task.

Click the Start button if you are ready to start Task 3.

Remember, you have **5 minutes** to solve this task.

Task 3



iv) Stage 4

Instructions - Task 4

For Task 4, you will be asked to make a series of **ten pairwise decisions** between **"Option A"** and **"Option B".**

If Task 4 is the one selected for payment, **your earnings depend on which of the 10 decisions is randomly selected**, you will be paid out according to this decision.

On the next page, you will find an example.

This is how the decision sheet will look like:

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

Now, please look at the **first decision:**

Option A has a 1 in 10 chance of paying **£2.20** and a 9 in 10 chance of paying **£1.75**

Option B has a 1 in 10 chance of paying **£4.20** and a 9 in 10 chance of paying **£0.10**

The other decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for the **last decision** in the bottom row, a random draw will not be needed since each option pays the highest payoff for sure, so your choice here is between £2.20 and £4.20.

To summarise, you will make ten choices: for each decision row, you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows.

Please click Continue if you have understood the instructions and wish to continue to the Decision table.

Click the Start button if you are ready to start Task 4.

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

i) **Final questionnaire part 1**

Please answer the following questions

Page 1/2

Question 1

On a scale from **0 (not at all like me)** to **10 (exactly like me),** please indicate the extent to which the following statement **describes you**.

Competition brings the best out of me.

[Options: 0 (not at all like me), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (exactly like me)]

Question 2

On a scale from **0 (completely unwilling to take risks)** to **10 (very willing to take risks)**, in general, how willing are you to **take risks**?

[Options: 0 (completely unwilling to take risks),1,2,3,4,5,6,7,8,9,10 (very willing to take risks)]

Question 3

On a scale from **1 (very poor)** to **7 (excellent),** how would you classify your maths skills?

[Options: 1(very poor),1,2,3,4,5,6, 7 (excellent)]

ii) Final questionnaire part 2

Please answer the following questions

Page 2/2

Question 1

How religious do you consider yourself to be?

[Options: Very religious, Moderately religious, Not religious]

Question 2

What is your political identity?
[Options: Strongly conservative, Somewhat conservative, Neutral, Somewhat liberal, Strongly liberal]

Question 3

What's your country/region of primary citizenship?

[Options: text input _____]

Question 4

What is your field of study?

[Options: text input _____]

Earnings

Please don't close this window.

Payment for **Task 1**: [Stage 1]

Payment for **Task 2**: [Stage 2]

Payment for **Task 3**: [Stage 3]

Payment for **Task 3**: [Stage 4]

The task chosen for payment is number: [Randomly chosen task]

This means that your total earnings are:

[£4 (participation fee) + Randomly chosen task + random bonus]

Please remain seated. The experimenter will soon come and pay you at your cubicle.

Thank you for taking part in this study.

Principals

[Please note that the instructions for the principals are the same in *GenderInfo* and in *NoGenderInfo,* with the exception of one screen. We point out this specific screen in the following.]

Introduction

Thank you for taking part in this study.

Please read the following information carefully.

If you have any questions, please let us know by raising your hand. Your question will then be answered in your cubicle.

You will receive **£4.00** for your participation.

The additional amount of money that you will earn **depends on your decisions and that of other players in the study.**

At the end of this study, you will receive the money that you have earned in cash.

The study consists of three parts, only two of which are eligible for payment. You will be paid only for one part, which will be selected at random at the end of the study. Each part has an equal chance of being selected.

Please do not close the last page of the study as it will determine your payment.

Each task will be introduced on your screen with the header **"Instructions".** The instructions will explain in detail what the respective part of the study is about.

One of the tasks involves summing numbers. You must solve the problems **without any helping device, i.e., devices such as pocket calculators or cell phones, are not allowed**. If you use any helping device, you will be immediately excluded from the experiment and from all payments. However, you are **allowed to use the pen and paper positioned next to your keyboard.**

Finally, please note that you are not allowed to communicate with other participants throughout the study and that unnecessary interference with computers will lead to exclusion from the experiment.

Question 1

What is your sex? [Options: Male, Female]

Question 2

How old are you? [Options: numeric input: _____]

Question 3

What is your country of residence? [Options: United Kingdom, Other]

i) **Stage 1**

Instructions - Trial

For this trial task, you will be asked to **calculate the sum of five randomly chosen numbers**.

Once you **click the Submit button**, the next numbers will appear.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

Your answers to the problems are anonymous.

You have **2 minutes** to solve as many as you can.

On the next page, you will find **an example.**

Instructions - Trial

Below you can see an **example** of the sums you will be asked to make.

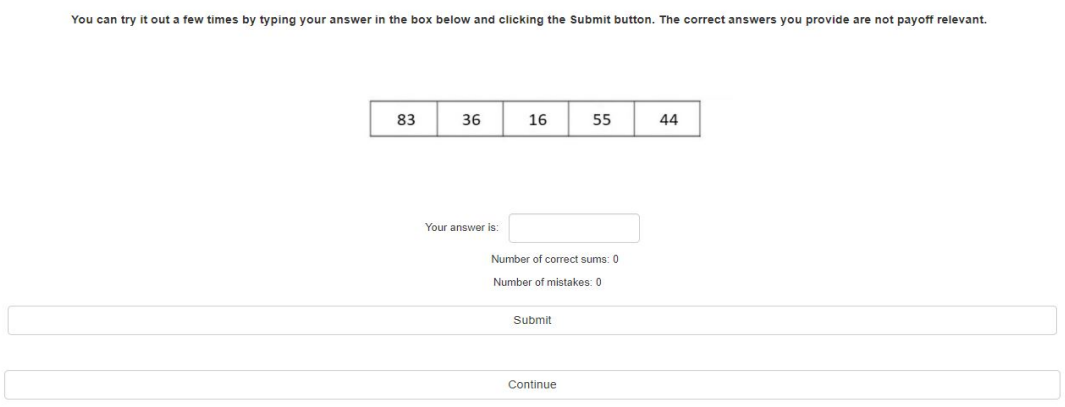
Your task is to add the numbers contained in the image:

83+36+16+55+44 = 234

You submit an answer by clicking the Submit button with your mouse.

When you enter an answer, the computer will immediately tell you whether your answer is correct or not.

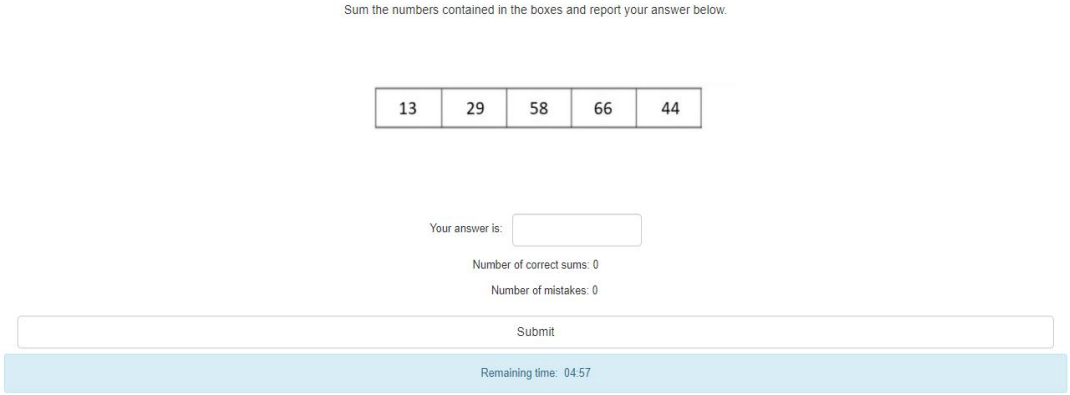
On your screen, you will also see the number of correctly solved sums and the remaining time.



Click the Start button if you are ready to start the trial.

Remember, you have **2 minutes**.

Trial



ii) <u>Stage 2</u>

<u>[Principals wait until agents finalise Stage 1 and Stage 2]</u>

iii) Stage 3 a

Instructions - Task 1

Instructions - Task 1

For this task, you and the other participants have been randomly assigned one of two roles, Player A or Player B.

You are Player B.

Your task is to choose **which payment scheme you want to apply to Player A's future performance.** That is, you need to choose which payment scheme you want the other player to work for in the next task, based on the information you will be provided.

You can choose between a **piece rate payment** or a **tournament payment** scheme. The information includes **Player A's past performance under the piece rate payment (Task 1) and the tournament payment scheme (Task 2)** but does not include their ranking compared to other players.

Player A will have **5 minutes** to solve correctly as many sums of a series of five numbers as possible. The task is the same as the one you undertook in the trial stage.

Instructions continue on the next page.

Instructions - Task 1

- If you choose the **piece rate payment**, both you and Player A will receive **£0.50 for each correct answer** provided by Player A.

Your earnings do not decrease if he/she provides an incorrect answer.

- If you choose the **tournament payment**, then **Player A's performance will be evaluated relative to the performance of his/her other three group members** in the Task 2-tournament.

If he/she correctly solves more problems than his/her group members did in Task 2, both you and Player A receive four times the payment from the piece rate, which is **£2 per correct problem**. However, you both receive **no earnings** for this task if Player A does not solve more problems correctly than his/her group members did in the Task-2 tournament. Player A is competing against the same group members as in Task 2 (2 males and 2 females). You will not be informed of how Player A did in the tournament until all four tasks have been completed.

If there are ties, the winner will be randomly determined.

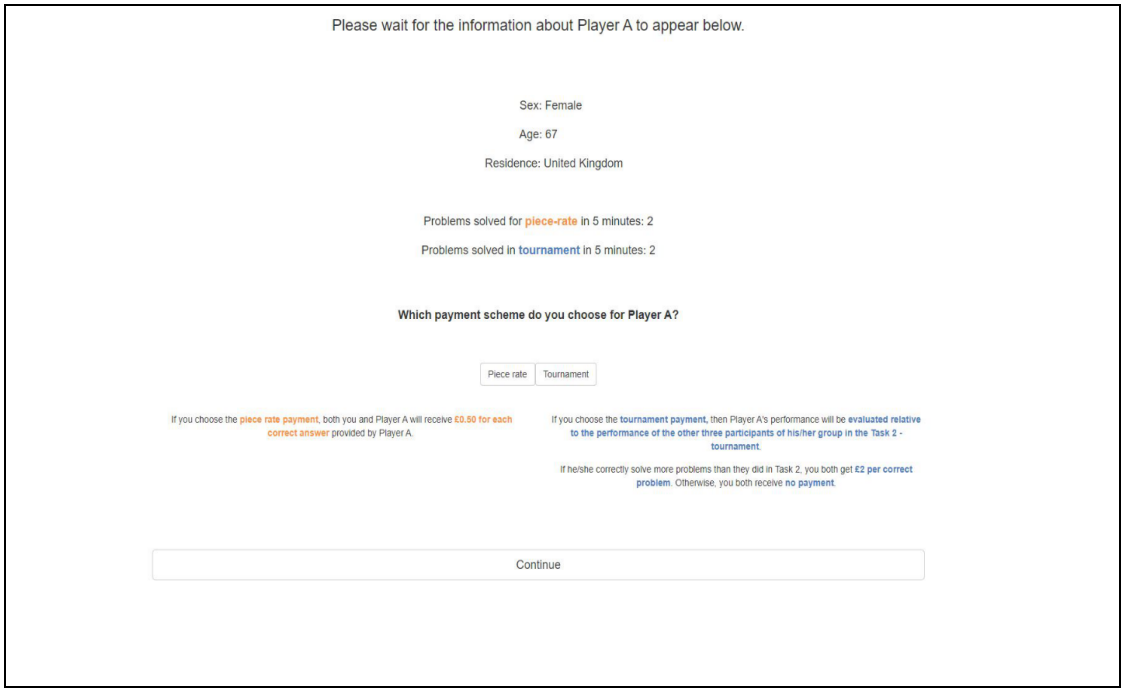
On the next page, you will find information about Player A.

Start of treatment-specific instructions

[On the next screens, the treatment variation is implemented. In *NoGenderInfo*, Player A's sex is not shown. Whereas in *GenderInfo*, the information is shown. We first present the treatment-specific instructions for *GenderInfo* before we present the treatment-specific instructions for *NoGenderInfo*. Recall that curly brackets indicate that one of the listed options was displayed to the principals.

[In *GenderInfo*, instructions include agent's sex and pronouns as follows:]

GenderInfo



Question 1

Bonus payment

You have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

Which payment scheme would Player A have chosen for {himself/herself}?

[Options: Piece rate, Tournament]

Question 2

If you had been assigned the role of Player A, which incentive scheme would you have chosen for **yourself**?

[Options: Piece rate, Tournament]

Question 3

Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

On a scale from **1 (Best 25%) to 4 (Worst 25%),** how well do you think Player A performed in the previous **piece-rate** stage relative to the other subjects **with the same role** in this session?

Remember, {he/she} solved [number of correctly solved problems in the piece rate stage] problems in the previous piece rate stage.

[Scale: 1, 2, 3, 4]

Question 4

Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

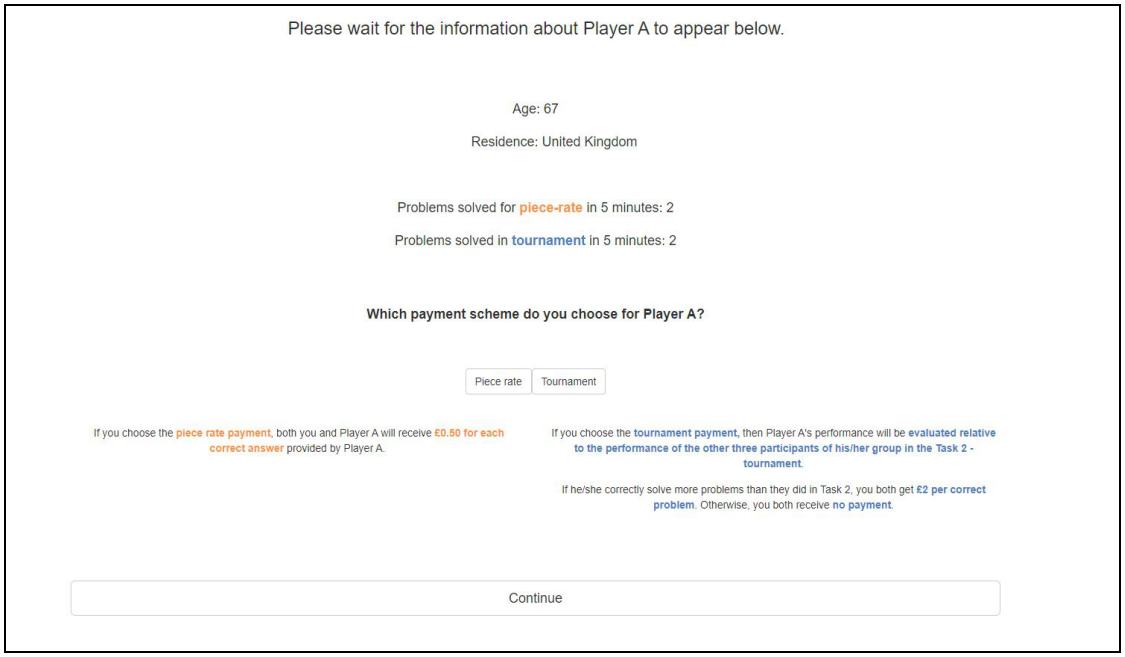
On a scale from **1 (Best) to 4 (Worst),** how well do you think Player A performed in the previous **tournament stage** relative to other subjects in **{his/her} group**?

Remember, {he/she} solved [Number of correctly solved problems in the tournament stage] problems in the previous tournament stage.

[Scale: 1, 2, 3, 4]

[In *NoGenderInfo*, instructions do not include agent's sex and pronouns as follows. Instead of using gender-specific pronouns, we refer to the agent as he/she, his/her and himself/herself. By contrast, in *GenderInfo*, we choose the pronoun that matches the gender of the agent.]

NoGenderInfo



Question 3

Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

On a scale from **1 (Best 25%) to 4 (Worst 25%),** how well do you think Player A performed in the previous **piece-rate** stage relative to the other subjects **with the same role** in this session?

Remember, he/she solved [number of correctly solved problems in the piece rate stage] problems in the previous piece rate stage.

[Scale: 1, 2, 3, 4]

Question 4

Bonus payment

Again, you have the chance to earn additional money. If this question is randomly chosen for payment, you will be paid an extra £0.50 on top of your final payoff if you answer it correctly.

On a scale from **1 (Best) to 4 (Worst),** how well do you think Player A performed in the previous **tournament stage** relative to other subjects in **his/her group**?

Remember, he/she solved [Number of correctly solved problems in the tournament stage] problems in the previous tournament stage.

[Scale: 1, 2, 3, 4]



End of treatment specific instructions

iv) Stage 3 b

We use the “Gender-Career” IAT that is made available on: Materials | iatgen (wordpress.com)

Instructions – Task 2

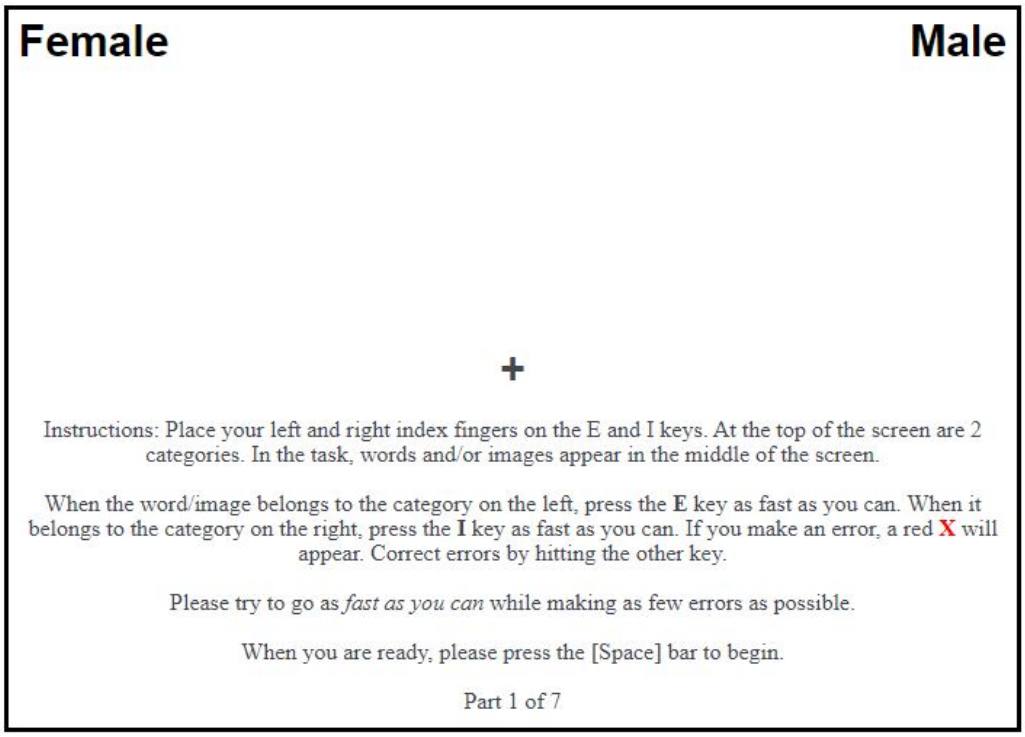
You will receive detailed instructions for this task on the next page.

You will be asked to insert the following random ID **before you can start the task**, please **make a note of it on a piece of paper** (as you will be asked for it on the next page).

random ID: [Random id]

After you finish the task, you will be given **another code.** You need to **copy and paste the code into the input box below the screen.**

This task is not payoff relevant.



v) **Stage 4**

Instructions - Task 3

For Task 3, you will be asked to make a series of **ten pairwise decisions** between **"Option A"** and **"Option B".**

If Task 3 is the one selected for payment, **your earnings depend on which of the 10 decisions is randomly selected**, you will be paid out according to this decision.

On the next page, you will find an example.

This is how the decision sheet will look like:

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

Now, please look at the **first decision:**

Option A has a 1 in 10 chance of paying **£2.20** and a 9 in 10 chance of paying **£1.75**

Option B has a 1 in 10 chance of paying **£4.20** and a 9 in 10 chance of paying **£0.10**

The other decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for the **last decision** in the bottom row, a random draw will not be needed since each option pays the highest payoff for sure, so your choice here is between £2.20 and £4.20.

To summarise, you will make ten choices: for each decision row, you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows.

Please click Continue if you have understood the instructions and wish to continue to the Decision table.

Click the Start button if you are ready to start Task 3.

	Option A	Option B
Decision 1	1/10 of £2.20 & 9/10 of £1.75	1/10 of £4.20 & 9/10 of £0.10
Decision 2	2/10 of £2.20 & 8/10 of £1.75	2/10 of £4.20 & 8/10 of £0.10
Decision 3	3/10 of £2.20 & 7/10 of £1.75	3/10 of £4.20 & 7/10 of £0.10
Decision 4	4/10 of £2.20 & 6/10 of £1.75	4/10 of £4.20 & 6/10 of £0.10
Decision 5	5/10 of £2.20 & 5/10 of £1.75	5/10 of £4.20 & 5/10 of £0.10
Decision 6	6/10 of £2.20 & 4/10 of £1.75	6/10 of £4.20 & 4/10 of £0.10
Decision 7	7/10 of £2.20 & 3/10 of £1.75	7/10 of £4.20 & 3/10 of £0.10
Decision 8	8/10 of £2.20 & 2/10 of £1.75	8/10 of £4.20 & 2/10 of £0.10
Decision 9	9/10 of £2.20 & 1/10 of £1.75	9/10 of £4.20 & 1/10 of £0.10
Decision 10	10/10 of £2.20 & 0/10 of £1.75	10/10 of £4.20 & 0/10 of £0.10

vi) **Final questionnaire part 1**

Please answer the following questions

Page 1/2

Question 1

On a scale from **0 (not at all like me)** to **10 (exactly like me),** please indicate the extent to which the following statement **describes you**.

Competition brings the best out of me.

[Options: 0 (not at all like me), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (exactly like me)]

Question 2

On a scale from **0 (completely unwilling to take risks)** to **10 (very willing to take risks)**, in general, how willing are you to **take risks**?

[Options: 0 (completely unwilling to take risks),1,2,3,4,5,6,7,8,9,10 (very willing to take risks)]

Question 3

On a scale from **1 (very poor)** to **7 (excellent),** how would you classify your maths skills?

[Options: 1(very poor),1,2,3,4,5,6, 7 (excellent)]

vii) Final questionnaire part 2

Please answer the following questions

Page 2/2

Question 1

How religious do you consider yourself to be?

[Options: Very religious, Moderately religious, Not religious]

Question 2

What is your political identity?
[Options: Strongly conservative, Somewhat conservative, Neutral, Somewhat liberal, Strongly liberal]

Question 3

What's your country/region of primary citizenship?

[Options: text input _____]

Question 4

What is your field of study?

[Options: text input _____]

Earnings

Please don't close this window.

Payment for **Task 1**: [Stage 3 earnings of agent]

Payment for **Task 3**: [Earnings in H&L]

The task chosen for payment is number: [Randomly chosen task]

This means that your total earnings are:

[£4 (participation fee) + Randomly chosen task + random bonus]

Please remain seated. The experimenters will call you one by one based on your computer number.

Thank you for taking part in this study.

Chapter IV - M/F treatment

When it comes to the treatments where the gender of the other player is known, I only report the part of the instructions that varies between them. The translations are available upon request.

Intro India

Please choose your preferred language.

कृ पयाअपनी पसंदीदाभाषा चुन।

English िहंदी

M/F English

Welcome

Thank you for taking part in this study.

In this study, you will be presented with information relevant to human decision making and asked to answer some questions about it. Please be assured that your responses will be kept completely confidential.

In total, the study should take you less than 10 minutes to complete.

On top of your participation fee, you can earn points depending on your choices and the choices of another participant in this study up to a maximum of 1000 points (1000 points = 220 ₹).

By clicking the Continue button below, you acknowledge that your participation in the study is voluntary, that you are 18 years of age or older, and that you are aware that you may choose to terminate your participation in the study at any time and for any reason, but you will not be paid if you do not finish it. That is, if you do close your browser window or leave the study, you will not be able to re-enter and we will not be able to pay you for your participation.

Instructions

You are Player 1 and, for the purpose of this study, you are going to be paired with another randomly drawn person, Player 2.

Shortly, we will give you more information about the personal characteristics of Player 2. Player 2 will be given the same information about your personal characteristics.

You and Player 2 have the opportunity to share 1500 points, but both of you have to agree on how to split them. You must vote for one of two options, Option A or Option B. Player 2 is asked to make the same decision.

If you and Player 2 vote for the same option, that option wins and is implemented.

If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

The two options are:

Option A



Option B



If you vote for Option A and Player 2 votes for Option B, you both earn 0 points.

If you vote for Option B and Player 2 votes for Option A, you both earn 0 points.

Your vote and the vote of Player 2 will be matched after you have taken your decisions.

You must vote for Option A or Option B without knowing what Player 2 has voted for. Similarly, Player 2 must vote without knowing what you have voted for.

Control questions

Quiz

You have a maximum of three attempts. If you fail to answer correctly the questions below more than three times, you will be removed from this study.

You are Player 1.



Remember, if you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

1) How many points are you going to get if both you and Player 2 vote for Option A?
2) How many points are you going to get if you vote for Option A and Player 2 votes for Option B?
3) How many points are you going to get if you vote for Option B and Player 2 votes for Option A?
4) How many points are you going to get if both you and Player 2 vote for Option B?

Remaining attempts to answer questions:

Error messages

You made at least one mistake.

Press Continue to try again.

You failed to provide the correct answer 3 times.

You will now be removed from the study.

info English

We will now randomly draw the person that will be matched with

you.

Player 2 is equally likely to be a man or a woman.

(Please wait for the information to appear below)

Player 2 is a man/woman from India (from a Hindi speaking state) who has been informed that you are a man/woman from India (from a Hindi speaking state).

Bonus Payment

You have the chance to earn additional money. If your answer to the question below is correct, you will be paid an extra 20 points on top of your final payoff.

Please give your best guess, which option is Player 2 going to vote for?

○ **Option A** (500 points for him/herself and 1000 points for me)

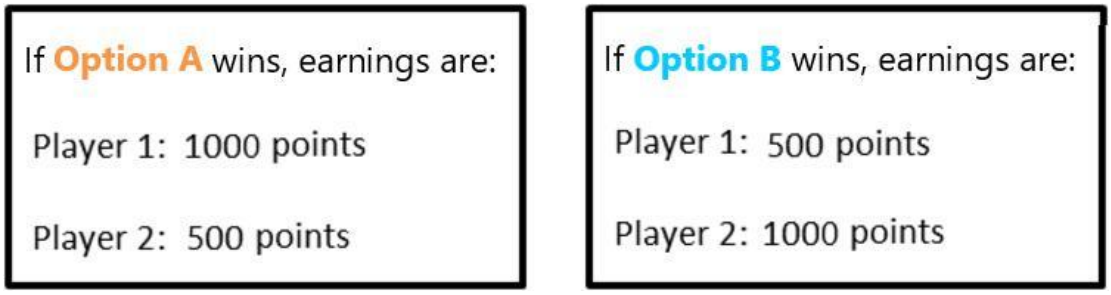
○ **Option B** (1000 points for him/herself and 500 points for me)

Voting decision

You are Player 1. Player 2 is a man/woman from India who has been informed that you are a man/woman from India.

You and Player 2 have the opportunity to share 1500 points.

Option A **Option B**



If you and Player 2 vote for the same option, that option wins and is implemented.

If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

Which option do you vote for?

Option A **Option B**

Part 2

In the next pages, you will be asked some questions related to your choice and the study in general. You do not earn any additional points in answering these questions. However, in order to finish you need to provide an answer.

1) You were asked to vote for Option A or Option B. Please write in the box below why you decided to vote for Option A

1) You were asked to vote for Option A or Option B. Please write in the box below why you decided to vote for Option B.

2) What do you think this study was about?

- ○ Selfishness
- ○ Competitiveness
- ○ Gender
- ○ None of the above

3) What do you think was the sex of Player 2?

- ○ Woman
- ○ Man

4) Take a moment to try and remember, what was the country of Player 2?

- ○ Nepal
- ○ India
- ○ Pakistan
- ○ None of the above

Chapter IV - FF treatment

We will now randomly draw the person that will be matched with you.

Player 2 is equally likely to be a man or a woman.

(Please wait for the information to appear below)

Player 2 is a woman from India (from a Hindi speaking state) who has been informed that you are a woman from India (from a Hindi speaking state).

Bonus Payment

You have the chance to earn additional money. If your answer to the question below is correct, you will be paid an extra 20 points on top of your final payoff.

Please give your best guess, which option is Player 2 going to vote for?

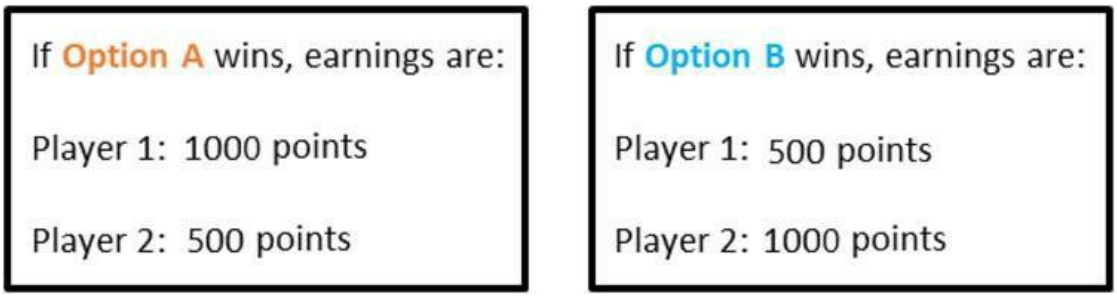
○ **Option A** (500 points for herself and 1000 points for me)
○ **Option B** (1000 points for herself and 500 points for me)

Voting decision

You are Player 1. Player 2 is a woman from India who has been informed that you are a woman from India.

You and Player 2 have the opportunity to share 1500 points.

Option A **Option B**



If you and Player 2 vote for the same option, that option wins and is implemented.
If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

Which option do you vote for?

Option A **Option B**

Chapter IV - FM treatment

We will now randomly draw the person that will be matched with you.

Player 2 is equally likely to be a man or a woman.

(Please wait for the information to appear below)

Player 2 is a man from India (from a Hindi speaking state) who has been informed that you are a woman from India (from a Hindi speaking state).

Bonus Payment

You have the chance to earn additional money. If your answer to the question below is correct, you will be paid an extra 20 points on top of your final payoff.

Please give your best guess, which option is Player 2 going to vote for?

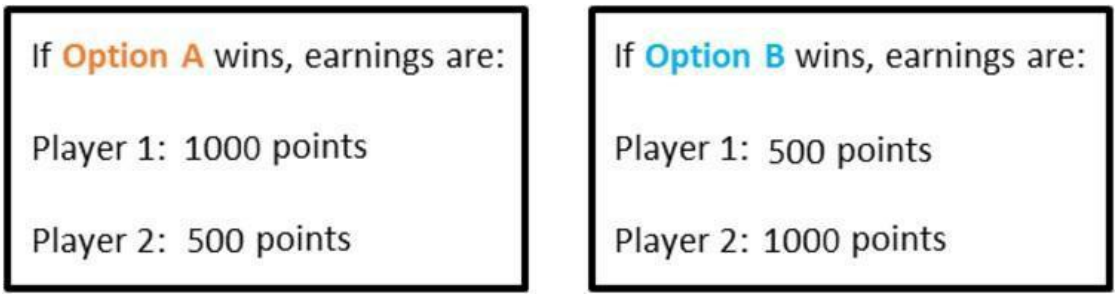
○ **Option A** (500 points for himself and 1000 points for me)
○ **Option B** (1000 points for himself and 500 points for me)

Voting decision

You are Player 1. Player 2 is a man from India who has been informed that you are a woman from India.

You and Player 2 have the opportunity to share 1500 points.

Option A **Option B**



If you and Player 2 vote for the same option, that option wins and is implemented.
If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

Which option do you vote for?

Option A **Option B**

Chapter IV - MM treatment

We will now randomly draw the person that will be matched with you.

Player 2 is equally likely to be a man or a woman.

(Please wait for the information to appear below)

Player 2 is a man from India (from a Hindi speaking state) who has been informed that you are a man from India (from a Hindi speaking state).

Bonus Payment

You have the chance to earn additional money. If your answer to the question below is correct, you will be paid an extra 20 points on top of your final payoff.

Please give your best guess, which option is Player 2 going to vote for?

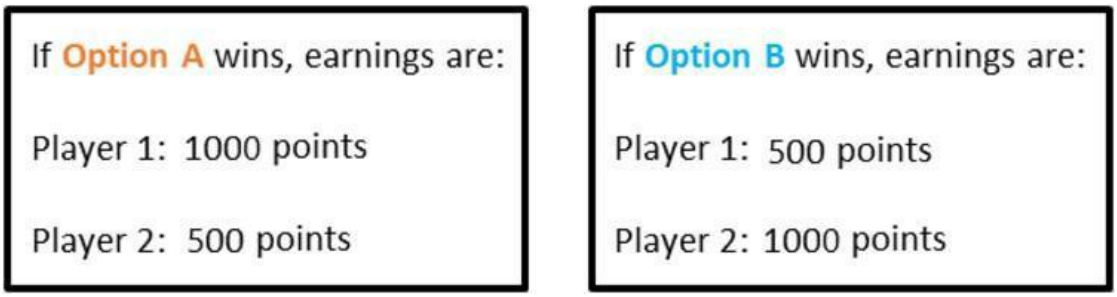
○ **Option A** (500 points for himself and 1000 points for me)
○ **Option B** (1000 points for himself and 500 points for me)

Voting decision

You are Player 1. Player 2 is a man from India who has been informed that you are a man from India.

You and Player 2 have the opportunity to share 1500 points.

Option A **Option B**



If you and Player 2 vote for the same option, that option wins and is implemented.
If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

Which option do you vote for?

Option A **Option B**

Chapter IV - MF treatment

We will now randomly draw the person that will be matched with you.

Player 2 is equally likely to be a man or a woman.

(Please wait for the information to appear below)

Player 2 is a woman from India (from a Hindi speaking state) who has been informed that you are a man from India (from a Hindi speaking state).

Bonus Payment

You have the chance to earn additional money. If your answer to the question below is correct, you will be paid an extra 20 points on top of your final payoff.

Please give your best guess, which option is Player 2 going to vote for?

○ **Option A** (500 points for herself and 1000 points for me)
○ **Option B** (1000 points for herself and 500 points for me)

Voting decision

You are Player 1. Player 2 is a woman from India who has been informed that you are a man from India.

You and Player 2 have the opportunity to share 1500 points.

Option A

If Option A wins, earnings are:

Player 1: 1000 points

Player 2: 500 points

Option B

If Option B wins, earnings are:

Player 1: 500 points

Player 2: 1000 points

If you and Player 2 vote for the same option, that option wins and is implemented.
If you and Player 2 vote for different options, no option wins and both you and Player 2 earn 0 points.

Which option do you vote for?

Option A **Option B**