Three essays on experimental economics

by

Amalia Rodríguez Valencia

Submitted in Partial Fulfillment
of the
Requirements for the Degree
Doctor of Philosophy

Supervised by
Professors Mariana Blanco and Darwin Cortés

Facultad de Economía
Doctorado en Economía
Universidad del Rosario

Bogotá - Colombia

2022

To Eliana, Margarita, Darío, Marta, and Gloria.
In loving memory of Ivonne and Eva.

Para Eliana, Margarita, Darío, Marta y Gloria.
En memoria de Ivonne y Eva.

Curriculum Vitae

The author was born in Medellín, Colombia. She attended Universidad de Antioquia, Medellín, from 2003 to 2008 and graduated first of her class with a Bachelor of Arts degree in Economics. In 2011, she moved to Toulouse, France, with a scholarship from Banco de la República, and received a Master of Sciences in Economics from the Toulouse School of Economics. In 2016 she began her PhD studies at Universidad del Rosario in Bogotá, Colombia. She was funded by the Fulbright Commission to spend a year as a Visiting Student Researcher at University of Maryland in College Park between 2018 and 2019.

Acknowledgments

I want to thank Mariana Blanco and Darwin Cortés for their tireless support and trust in me and my work, even when it was hard for me to believe in myself. Thank you for understanding my process and enabling the conditions that made it possible to move forward with my research, even when a pandemic got in the way of my plans. I have grown immensely thanks to your mentorship, and although it may not always seem so, I have always listened.

I also thank Çağatay Kayı for always being attentive to my process and Diego Aycinena and José Guerra, from whom I learned so much as a research assistant. To César Mantilla, Santiago Sautua, Catalina Franco, Francesco Bogliacino, and other regular participants of the Bogotá Experimental and Behavioral Economics Seminar (Universidad de los Andes and Universidad del Rosario), I gratefully acknowledge all the feedback that helped improve my work over the years. You created a space in which I always felt comfortable presenting my progress with the confidence that all comments would only be constructive. A big thank you also to Patricia Peña, Angélica Velandia, and Ana María Maldonado for all their assistance over the years.

I am also grateful to my fellow students at Universidad del Rosario and to the friends I made during my academic visit to the University of Maryland, from whom I learned so much. And to friends from before and those I met during this time, thank you for always being willing to listen to my projects and read my drafts: Marine Hainguerlot, Camila Galindo, Jorge Flórez, Sebastián Ramírez, Vanessa Ospina, Diana Restrepo, Yuliet Verbel, Amma Panin, and Cristina Bratu. Your friendship and encouragement have been invaluable during this time, which has been challenging in more ways than one. To Sara Atehortúa, my friend and co-author, thank you for motivating me in the hardest times and for making me part of your family.

And above all, thanks to my family. Papás, Eli, tías, abuela, Eva y Dani, han sido mi soporte y a la vez mis alas. Gracias por apoyar mis proyectos y por ser el lugar al que siempre quiero regresar.

Resumen

Esta tesis doctoral está compuesta por tres capítulos. En el capítulo 1, junto con Sara Atehortúa, proponemos un experimento de laboratorio para entender si entornos de restricción de recursos e incertidumbre sobre las necesidades relativas de potenciales beneficiarios afectan las decisiones de asignación de los médicos, y cómo ocurre. Cuando hay incentivos para la sobre provisión, encontramos que un paciente atendido por un médico con restricciones y bajo incertidumbre obtiene mayores beneficios y recibe asignaciones más cercanas a su óptimo que los pacientes vistos por médicos sin restricciones o que deciden sólo bajo incertidumbre. Además, observamos una redistribución de recursos cuando los médicos deciden bajo restricciones de recursos e incertidumbre. En particular, cuando los recursos son escasos, los médicos tienden a asignar los servicios limitados a los pacientes con mejor estado inicial de salud en ausencia de tratamiento, con una mayor capacidad para beneficiarse de recursos adicionales, con baja necesidad de servicios en el óptimo, y a quienes alcanzan el menor beneficio máximo posible. Por último, encontramos que las restricciones, con o sin incertidumbre, llevan a los médicos egoístas a aproximarse a lo que es mejor para el paciente.

En el capítulo 2, en coautoría con Mariana Blanco y Darwin Cortés, proponemos un experimento de laboratorio para entender cómo la exclusión social afecta a los participantes de esta interacción antisocial en términos de rendimiento y emociones reportadas. Adoptamos una manipulación de ostracismo ampliamente utilizada en psicología y la llevamos a un laboratorio de economía experimental. Encontramos que los eventos de exclusión social sólo afectan a los participantes ligeramente excluidos y que este efecto se explica por las emociones generadas tras la exclusión. Además, las víctimas de la exclusión informan de reducciones en la dimensión de valencia de las emociones, al igual que los que tienen la opción de excluir pero deciden no hacerlo. La posibilidad de que los espectadores castiguen a los posibles victimarios, la desaprobación generalizada de los eventos de exclusión y una aprobación extendida de la inclusión reducen la incidencia de este fenómeno. Esta reducción se produce a costa de cambios negativos en las emociones declaradas por la mayoría de los participantes, pero no se traduce en cambios en el rendimiento en la tarea. Por último, observamos que la exclusión previa aumenta la posibilidad de que las antiguas víctimas decidan excluir cuando tienen esa posibilidad.

En el capítulo 3, propongo un diseño experimental para analizar cómo tener líderes mujeres y hombres favorece la aceptación de una norma comunal sobre una norma agéntica en un entorno laboral. Además, estudio cómo la composición de género de un grupo y la elección de un líder por mérito favorecen la adopción de una distribución igualitaria. Encuentro que, aunque su preferencia individual es contraria a una norma igualitaria, los líderes masculinos son más influyentes que los femeninos a la hora de favorecer los cambios hacia la igualdad. Además, ninguna composición mixta de género en particular favorece una norma sobre la otra. Sin embargo, las mujeres de los grupos con mayoría femenina son menos propensas a seleccionar una compensación de torneo y a cambiar a una norma igualitaria cuando forman parte de un grupo.

Summary

This Ph.D. thesis consists of three chapters. In chapter 1, with Sara Atehortúa, we propose a lab experiment to understand if environments of resource restrictions and uncertainty on the relative needs of future beneficiaries affect physicians' resource allocation decisions and how. When there are incentives to over-treat, we find that a patient tended by a constrained physician under uncertainty obtains higher benefits and receives allocations closer to her optimum than patients from physicians with no constraints or deciding under uncertainty alone. In addition, we observe a redistribution of resources when physicians decide with resource restrictions and uncertainty. In particular, when resources are scarce, physicians tend to allocate limited services to patients with higher benefits in the absence of medical services, a higher capacity to benefit from the resources, the scantiest need for service units, and the lowest benefits at the optimum. Finally, we find that constraints, with or without complete information on patient characteristics, lead selfish physicians to approximate what is best for the patient.

In chapter 2, in co-authorship with Mariana Blanco and Darwin Cortés, we propose a laboratory experiment to understand how social exclusion affects the participants of this antisocial interaction in terms of performance and reported emotions. We adopt a widely used ostracism manipulation from psychology and take it to an experimental economics laboratory. We find that social exclusion events only affect lightly excluded participants and that this effect is explained by the emotions generated after exclusion. In addition, the victims of exclusion report reductions in the valence dimension of emotions, as do those who have the option to exclude but decide not to. The possibility of bystanders punishing potential offenders, the generalized disapproval of exclusion, and an extended approval of inclusion reduce the incidence of exclusion. This reduction comes at the cost of negative changes in the reported emotions of most participant types, but it does not translate into changes in performance in a task. Last, we find that previous exclusion increases the decision of former victims to ostracize.

In chapter 3, I propose an experimental design to analyze how female or male leaders favor the adoption of a communal norm over an agentic one in a workplace setting. I also study how the gender composition of a group and the selection of a leader via merit favor the adoption of an egalitarian distribution. Although their individual preference

is against an egalitarian rule, male leaders are more influential than female leaders in favoring changes toward equality. In addition, no particular mixed-gender composition favors one norm over the other. However, women in female-majority groups are less likely to select a tournament compensation and to change to an egalitarian rule when they are part of a group.

Contents

Dedication ii
Curriculum Vitae iii
Acknowledgments iv
Resumen v
Summary vii
List of figures xii
List of tables xv
Foreword 1

1 Physician's Allocation Preferences under Scarcity and Uncertainty **2**
1.1 Introduction 2
1.2 Experimental design and procedures 5
1.2.1 Experimental design 5
1.2.2 Experimental procedures 10
1.3 Hypotheses 11
1.4 Results 13
1.4.1 Do budget constraints and uncertainty affect allocation decisions by physicians? 13
1.4.2 How does this happen? 17
1.5 Discussion and conclusions 23

2 Social exclusion in the lab **25**
2.1 Introduction 25
2.2 Experimental Design and procedures 28
2.3 Hypotheses 33
2.4 Results 34
2.4.1 Social Exclusion 36

2.4.2 Punishment 45
2.4.3 Role Reversal 55
2.5 Discussion and conclusion 58

3 Leaders and communal norms in the workplace 60
3.1 Introduction 60
3.2 Experimental design and procedures 64
3.2.1 Experimental design 64
3.2.2 Belief elicitation tasks 68
3.2.3 Experimental procedures 68
3.3 Results 69
3.3.1 Individual decisions 70
3.3.2 Group decisions 78
3.3.3 The role of leaders 88
3.4 Discussion and conclusions 96

Appendices 108

A Chapter 1 109

B Chapter 2 114

C Chapter 3 120

List of Figures

1.1 Profiles of patient health benefits (EMUs) 6
1.2 Medical services allocated by profile. 14
1.3 Distribution of physician types. 20

2.1 Group Stage from player A's point of view 30
2.2 Experimental Design 31
2.3 Average changes in normalized RET performance by exclusion category, by role. 37
2.4 Two dimensions of the reported emotions by intensity of the exclusion, by role 41
2.5 Average social exclusion with and without punishment, by round 45
2.6 Potential offender's behavior depending on previous ratings 46
2.7 Average changes in normalized RET by the reactions to social exclusion decisions, by role. 49
2.8 Changes in the two dimensions of the reported emotions by punishment, by role 51

3.1 Timeline of the experiment 64
3.2 Normal CDF of performance in encryption task, by sex 72
3.3 Proportion of participants that choose tournament by performance quartile, by gender 73
3.4 Proportion of participants that choose tournament by belief of position in group, by gender 76
3.5 Proportion of Participants that Choose Tournament by Performance Quartile, by Gender 82
3.6 Proportion of Participants that Choose Tournament by Performance Quartile, by Gender and Group Composition 82

3.7 Proportion of Participants that Choose Tournament in Stage 5 by Beliefs, by Gender 84

A.1 Group Screen - Control 109
A.2 Patient Screen - Control 110
A.3 Group Screen - BC Treatment 111
A.4 Group Screen - U Treatment 111
A.5 Patient Screen - BC+U Treatment 112
A.6 Lottery to measure risk aversion preference 112
A.7 Lotteries to measure prudence preference 113

B.1 Individual stage RET screen 114
B.2 Images from the Emotion Test 114

C.1 Experimental screen: encryption task 120

List of Tables

1.1 A physician's profits, revenues and costs by number of medical services allocated to a patient (EMUs). 7
1.2 Treatments. 9
1.3 Medical services allocated, percentage deviation from patient's optimal level, and patient health benefits. 14
1.4 Within-subjects analysis. 16
1.5 Patient's characteristics (within-subjects analysis) 18
1.6 Physician's type (within-subjects analysis) 22

2.1 Summary statistics and mean comparison tests, by role 36
2.2 OLS results for within-subject changes in normalized RET performance by exclusion, for potential offenders and victims 38
2.3 OLS results for within-subject changes in normalized RET performance by exclusion, pooled bystanders 40
2.4 OLS results for within-subject changes in reported emotions, for potential offenders and victims 43
2.5 OLS results for within-subject changes in reported emotions, for bystanders 44
2.6 Logit results for exclusion behavior after feedback 47
2.7 OLS results for within-subject changes in RET performance by punishment, for potential offenders and victims 50
2.8 OLS results for within-subject changes in reported emotions by punishment, for potential offenders and victims 53
2.9 OLS results for within-subject changes in reported emotions by punishment, for bystanders 54

2.10 Exclusion behavior of former potential victims in the role of potential offender 55
2.11 OLS results for within-subject changes in RET, role reversal 57

3.1 Alternatives per stage 66
3.2 Summary statistics and mean comparison tests, by treatment 70
3.3 OLS regressions of decision to select a tournament compensation, with and without externality 74
3.4 Ordered probit of belief of relative position 75
3.5 OLS regressions of decision to select a tournament compensation, with and without externality 77
3.6 OLS regressions of decision to select a tournament compensation with externalities, in groups 79
3.7 OLS regressions of decision to select a tournament compensation in groups when individually selecting not to 80
3.8 OLS regressions of beliefs of individual and group's relative position . . 83
3.9 OLS regressions of decision to select a tournament compensation with information of group composition 85
3.10 OLS regressions of decision to select a tournament compensation with information of group composition 86
3.11 OLS regressions of decision to change from an egalitarian alternative to tournament compensation 87
3.12 OLS regressions of decision of leader to select a tournament compensation (treatment) 89
3.13 OLS regressions of decision of leader to select a tournament compensation (gender composition of groups) 90
3.14 OLS regressions of decision of non-leaders to accept egalitarian distribution 92
3.15 OLS regressions of decision of non-leaders to only accept egalitarian distribution, including beliefs of leader's preferences 93
3.16 OLS regressions of decision of non-leaders who prefer Alternative 2 to accept egalitarian distribution 95

A.1 Physicians by configuration 110
A.2 Summary statistics and mean comparison tests, by treatment 113

B.1 OLS results for within-subject changes in normalized RET performance by exclusion, by role. Detailed table 115
B.2 OLS results for within-subject changes in reported valence by exclusion, by role. Detailed table 116
B.3 OLS results for within-subject changes in reported arousal by exclusion, by role. Detailed table 117
B.4 Punishment behavior of bystanders, with and without role reversal . . . 118
B.5 OLS results for within-subject changes in RET performance by punishment, for bystanders 119

C.1 OLS regressions of decision to select a tournament compensation, with and without externality. Top performing participants 121
C.2 OLS regressions of decision to select a tournament compensation, with and without externality. Belief of top position. 122
C.3 OLS regressions of decision to select a tournament compensation with information on group composition 123
C.4 OLS regressions of decision to select a tournament compensation with information on group composition. Top performing participants 124
C.5 OLS regressions of decision to select a tournament compensation with information of group composition 125
C.6 Summary statistics and mean comparison tests, by treatment 126

Foreword

This thesis is a collection of joint and solo essays on experimental economics. First chapter is a joint work with Sara Atehortúa. Second chapter is a joint work with Mariana Blanco and Darwin Cortés. Third chapter is a solo essay and my job market paper.

Chapter 1

Physician's Allocation Preferences under Scarcity and Uncertainty

1.1 Introduction

The COVID-19 contingency has forced doctors worldwide to prioritize intensive care units, indirectly determining who lives or who dies. More recently, governments have put a requirement in place to schedule the administration of an insufficient number of vaccines. This situation has unveiled a decision problem that was already present in other medical settings, such as the battlefield and rural doctors, general practitioners acting as gatekeepers to the healthcare system, and emergency room physicians with infrastructure limitations. In such settings, a physician must decide on the sequential allocation of the limited resources in a context in which, when facing a patient, she is unaware of the needs of potential future beneficiaries in the same resource pool. These decision-making problems are not alien to other contexts. For example, altruistic or inequality-averse resource managers in a company, the public sector, or a university also have to sequentially decide how to distribute their department's budget during a given period.

We propose a lab experiment to understand if resource constraints and uncertainty (on potential beneficiaries' relative needs) affect resource allocation in a medical context. If so, how these conditions interact with physicians' types and patients' characteristics, and if they induce redistributing effects that determine which beneficiaries achieve higher levels of well-being. In our experimental design, all participants decide as physicians on the number of medical service units to allocate to seven different patient profiles. For each profile, it is possible to identify the efficiency of the allocations since the optimal provision is known. We randomly assign university students connected virtually to the sessions to three alternative treatments. In the Budget Constraint (BC) treatment, subjects face a restriction on the total number of medical service units they can allocate to a group of three patients. In the Uncertainty (U) treatment, we include uncertainty on the profile of the last patient of the group. In the Budget Constraint and Uncertainty (BC+U) treatment, subjects face uncertainty and resource restrictions simultaneously. Subjects participate in a control treatment without constraints or uncertainty before the treatment round. Our setup follows the basic features of Hennig-Schmidt et al. (2011) and incorporates elements from Brosig-Koch et al. (2016) and Martinsson and Persson (2019).

We find that when there are no resource constraints or uncertainty, the incentives to treat translate into over-provision of resources to all patient profiles. On the contrary, there is under-provision when subjects face budget constraints (with or without uncertainty). Interestingly, in absolute terms, under-provision is smaller that over-provision, which means that patients are more distant from their optimal allocation when they are over-provided. We also find that a combination of budget constraint and uncertainty increases patients' benefits compared to when physicians are under uncertainty alone. In addition, these conditions induce within-group redistribution benefiting patients with better initial health conditions, greater capacity to benefit from additional medical services, lower optimal needs, and those whose potential maximum benefit is the smallest among all.

In the presence of constraints with or without uncertainty, we observe that only selfish physicians, as categorized using a prioritization criteria, change their behavior, granting higher benefits for patients. This finding is of particular interest because it can justify some budget constraints when there are incentives to over-treat in an uncertain environment. As a result, when physicians are unaware of future patients' relative needs, using constraints as an expense-containment strategy can generally improve patient outcomes, compared to when there are no restrictions.

Experimental health economics studying allocation decisions is extensive. From the seminal paper of Hennig-Schmidt et al. (2011), a series of works have emerged to answer the empirical question of how payment schemes affect physician's provision behavior (Brosig-Koch et al., 2013, 2016, 2017, 2019; Keser et al., 2014, 2020). In these studies, students in the lab face alternative incentives for medical service assignments in a setting free from constraints.[1] Resource constraints also appear concurrently as a critical feature in the allocation decisions research. One branch of such literature examines the distribution principles that may drive these decisions when resources are limited (Ahlert et al., 2012; Ahlert and Funke, 2012; Ahlert et al., 2013; Ahlert and Schwettmann, 2017). These works share a similar setup that they implement with students (medical or others), using a medically or neutrally framed experiment. Other studies analyze the performance of different payment schemes under resource restrictions (Di Guida et al., 2019; Oxholm et al., 2019, 2021).

Harnessing the versatility of the experimental setup from Hennig-Schmidt et al. (2011) and Brock et al. (2016), some recent papers include uncertainty on the patient's health benefit (Martinsson and Persson, 2019; Hafner et al., 2017) or budget restrictions on the number of medical units to allocate to patients with known characteristics (Brendel et al., 2021). We contribute to this literature of resource restrictions and allocation decisions in two ways. First, we introduce uncertainty in the subject pool's characteristics in an environment with scarcity. We acknowledge uncertainty only in its risk dimension since physcians are unaware of the relative needs future beneficiaries of the scarce resources may have. We argue that physicians accurately estimate the probability distribution of the patients' health status since they are aware of the epidemiological prevalence and incidence of diseases. Thus, we are not interested in ambiguity. Second, different from Brendel et al. (2021), we assume that physicians have incentives to assign medical services to patients.

The rest of the paper is organized as follows. In Section 2, we describe the experimental design and procedures. Then, in Section 3 we move to detail our hypotheses. In Section 4, we present the results. In Section 5, we discuss results and breakdown our conclusions.

[1] Other studies dealing with payment schemes in a setting without constraint but using a different experimental design are Green (2014), Bejarano et al. (2017), and Lagarde and Blaauw (2017).

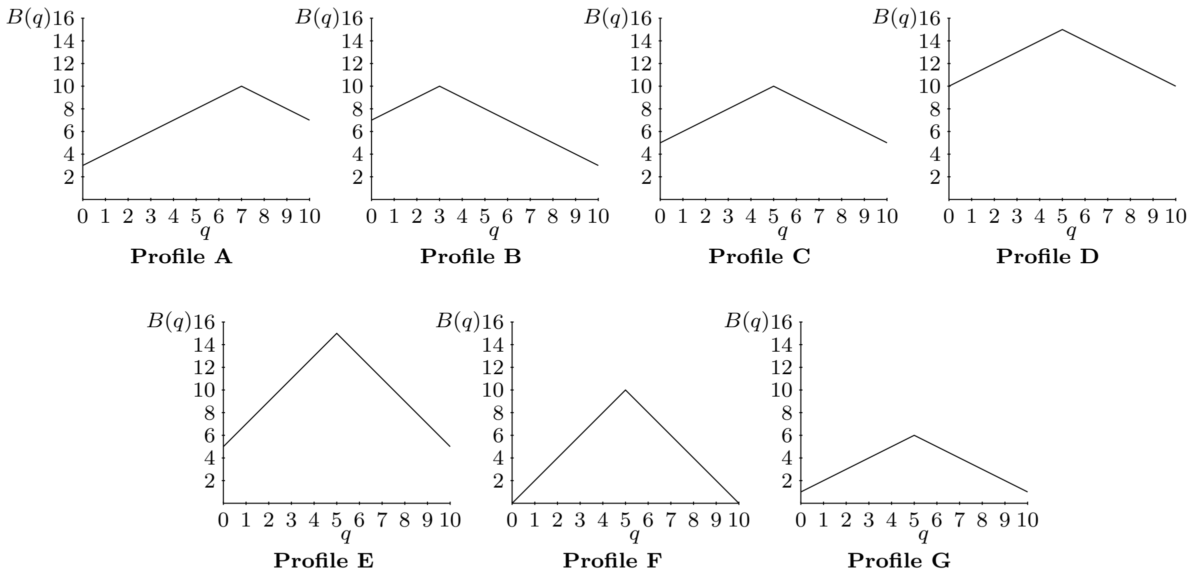
1.2 Experimental design and procedures

1.2.1 Experimental design

Participants decide as physicians on the number of medical service units $q \in [0, 10]$ to provide to seven patient profiles (A to F as seen in Figure 1.1). The number of service units defines each patient's profile health benefits $B(q)$ (see Figure 1.1). If, for example, a patient with profile A receives five units of medical services from a physician, she will procure a health benefit of 8 Experimental Currency Units -EMUs-.[2] We will use the terms patient and profile interchangeably from here onward. The seven profiles differ in four dimensions (see Figure 1.1): the optimal amount of medical services $\hat{q}$ (either 3, 5, or 7), that maximizes $B(q)$ in the y-axis (either 6, 10, or 15), the capacity to benefit from each additional unit of service (marginal benefit of either 1 or 2), and patient's initial health status in the absence of medical care $B(0)$ (either 0, 1, 3, 5, 7, or 10). Note that the benefit function is symmetric around the optimum (Ellis and McGuire, 1986), meaning that over and under provision is equally damaging for patients. Since the optimal amount of treatment units is known, there is a reference point to define under-provision and over-provision of the services (efficiency of the allocation decisions).

[2]EMUs are converted into COP at a rate of 1:0.8.

Figure 1.1: Profiles of patient health benefits (EMUs)



Notes: These figures depict patients' health benefits in EMUs (on the y-axis) over the number of medical service units (on the x-axis). Profiles A, B, and C come from Brosig-Koch et al. (2016), and E from Martinsson and Persson (2019). We introduce three additional patient profiles: D, F, and G.

The physician faces a trade-off when she decides on the units to allocate to the patient between her profit $\Pi(q)$ and the patient's health benefit.[3] The physician's profit is $\Pi(q) = R(q) - c(q)$ (see Table 1.1 for details on this function and parametrization). We argue that physicians have conflicting incentives as they aspire to signal ability to patients, families, and others by allocating resources (positive revenue) while also facing pressures to ration these resources from employers, colleagues, government, or insurers (cost) (Dusheiko et al., 2006; Brock et al., 2016; Godager et al., 2016). Both the revenue and the cost increase on the number of medical services assigned, and, for simplicity, we assume physicians have a homogeneous technology.

[3]This idea was first introduced by Ellis and McGuire (1986), where they define a cost-sharing model in which physicians decide on the level of services and consider both the patient's benefit and the hospital profits. Choné and Ma (2011) propose an adaptation of the model in which the physician optimizes a utility function that depends on both her profit and the patient's health benefit. We follow the latter approach here.

Table 1.1: A physician's profits, revenues and costs by number of medical services allocated to a patient (EMUs).

	Number of medical services (q)										
	0	1	2	3	4	5	6	7	8	9	10
$R(q)$	0	2	4	6	8	10	12	14	16	18	20
$c(q)$	0	0.1	0.4	0.9	1.6	2.5	3.6	4.9	6.4	8.1	10
$\Pi(q)$	0	1.9	3.6	5.1	6.4	7.5	8.4	9.1	9.6	9.9	10

Notes: $R(q) = pq$ is the revenue, where p is a fee per service (assumed as 2 following Brosig-Koch et al. (2016)). The cost $c(q) = \beta q^2$, with $\beta = 0.1$ following Hennig-Schmidt et al. (2011). We take these parameters intending to converse with the literature.

Our experiment has two parts. In Part 1, we arrange the seven profiles into three groups of three patients. We have two alternative configurations: ABD, DGC, FEG, or ABC, DGF, FED.[4] This means that a physician who faces the first configuration initially sees profiles A, B, and D. Alternatively, a physician who faces the second configuration meets a group consisting of patients A, B, and C. We randomly assign the configurations at the individual level and change profiles' labels to avoid subjects' recollection of the group configurations.[5] Here, subjects have 30 units of medical services available to allocate to the three patients from each group (control condition). At the top of the screen, participants can observe the 30 units they have available for allocating to the three patients in each group (see Figure A.1, in the Appendix). This means that a physician can potentially obtain her maximum profit level for every patient and that the optimal number of medical services is always achievable. The information on group composition and order of the patients is complete, and decisions are sequential. It is important to note that once they decide, the previously allocated services get discounted from the medical services' total budget.

In Part 2, we include two possible conditions: budget constraint and uncertainty. Budget constraint in our setting means that the physician has only ten (and not 30)

[4]Note that the profiles do not repeat within a group but can appear in more than one group.
[5]Participants observe labels F instead of A and vice versa and C instead of E and vice versa.

medical service units to distribute to each group of three patients.[6] We decided on the ten-unit benchmark for two main reasons. First, we need a constraint that separates selfish physicians from those with an equal split rule.[7] Second, we want to make the constraint salient so that the physician cannot allocate all three patients with their optimal, and, as a result, it requires her to make distributional decisions. Uncertainty means that each physician will encounter three groups with four (instead of three) patients each. A physician under uncertainty has complete certainty over the first two patients and is aware of the order in which they will arrive. However, the third and last patient is one of two alternative profiles, with a 50% probability.[8] We include uncertainty on the third patient, instead of the pool of patients, as we are interested in the medical decision to the current patient, with known characteristics, when the prospective patients' features are unknown. Furthermore, we focus on the risk dimension of uncertainty as we argue that physicians have a prior on the epidemiological distribution of the patients they will face, limiting ambiguity.[9]

In this part, we randomly assign participants at the session level to one of three treatments that use one or two of the described conditions: Budget Constraint -BC-, Uncertainty -U-,[10] and Budget Constraint and Uncertainty -BC+U- (see Table 1.2). In BC, subjects have complete information on the patients, and have up to ten service units to allocate to the three patients in each group (see Figure A.3, in the Appendix). In U and BC+U, subjects must decide allocations to patients in an uncertain environment. The critical difference between U and BC+U is the number of medical service units available for allocation. Subjects in treatment U have 30 medical service units

[6] As before, subjects observe one of the two alternative configurations. This means that subjects who faced ABD, DGC, FEG, in Part 1, can either face ABD, DGC, FEG, or ABC, DGF, FED in Part 2, resulting in 4 types of subjects (depending on the combination of both configurations in both parts of the experiment) (see Table A.1, in the Appendix).

[7] This principle rules out all multiples of three as a selfish physician would maximize her profit when she equally splits treatment units.

[8] All physicians have information that the first group of patients will be AB C/D. This means that patient A will arrive first, and B will arrive second; however, the third and last patient can be C or D with a 50% probability. The same follows for groups 2 (DG C/F) and 3 (FE G/D) (see Table A.1, in the Appendix).

[9] In this context, two patients are enough for studying the allocation decisions under risk. However, as we are also interested in the distributional principles that arise in the presence of a restriction, we require to have at least three patients.

[10] We include this treatment to ensure completeness in our design and to be able to separate the effects of resource restriction and uncertainty in situations such as those that motivate us.

to allocate to the three patients in each group. Subjects in the BC+U treatment have ten medical service units to distribute within each group (see Figures A.4 and A.5, in the Appendix). As in Part 1, we discount the units allocated to the total budget after each decision. In the BC and BC+U treatments, this limits the number of units they can distribute to future patients.

Table 1.2: Treatments.

	Constraint	
	Not binding	Binding
No uncertainty	Control	BC
Uncertainty	U	BC+U

Notes: All subjects play Part 1. In Part 2 they are randomly allocated (at the session level) to one of three treatments: Budget Constraint -BC-, Uncertainty -U-, and Budget Constraint and Uncertainty -BC+U-.

Although all subjects decide as physicians, we randomly match participants with three other subjects connected to the same session for payment purposes (payment-group). The computer randomly selects one of the groups of three patients from Part 1 and one from Part 2 to determine participants' earnings. It assigns payment-group members to a role: Physician, Patient 1, Patient 2, or Patient 3. The decisions of participants in the role of physician determine the payments for all group members.[11] Subjects in a physician's role obtain the monetary equivalent of the sum of the profits corresponding to her allocation to the three patients of her group. Subjects in each patient's role receive the monetary equivalent of the health benefit that matches her

[11]With this design feature, we depart from what is standard in the literature that uses similar designs. Typically, subjects in the physician's role decide how much to allocate to potential patients, and these allocation decisions become transfers to medical-related institutions or charities. Since we are interested in understanding distributional decisions, having a single final beneficiary does not make sense in our context. An alternative can be to select several potential recipients. However, in this case, subjects' preferences for a particular recipient can distort her distributional decisions.

group's physician decision. The composition of the payment-group is entirely anonymous for everyone, and the roles assigned can change between parts.

1.2.2 Experimental procedures

We carried out the experiments in nine online sessions (3 per treatment) with students from Universidad del Rosario between June 26 and July 14 of 2020, during a mandatory lockdown from the Colombian government.[12] We invited participants to an activity that could last for 2 hours, including an identity check-up, and asked them to be available 5 minutes before the beginning of the session. The experiments were programmed in z-Tree (Fischbacher, 2007) and conducted using z-Tree unleashed (Duch et al., 2020). We used the Online Recruitment System for Economic Experiments (ORSEE) (Greiner et al., 2004) from Rosario Experimental and Behavioral Economics Lab (REBEL) to recruit the subjects and the Zoom platform for the experimental sessions. We admitted participants to the session from a waiting room one by one. Once in the session, we changed their screen name for an id they should use for all the following experimental procedures. We randomly selected four subjects from each session for identity verification; for this, we used Zoom's private rooms. Only in one case, we could not verify the subject's identity; we excluded this participant from the experiment before the beginning of the session and replaced her with another subject.

At the beginning of each session, we read aloud both the general and Part 1's instructions to all the participants via Zoom. The remaining instructions were available on the experimental screens. The chat was open for questions throughout the session.[13] Before the decision round, we presented participants with the group of patients they were to face (Group Screen). We detailed the health benefit for each patient in the group and at every medical service level (see Figures A.1 and A.3, in the Appendix). Participants decided at their own pace when to begin with the decision rounds. Once they moved from the Group Screen, they faced each patient's screen. A Patient Screen

[12]Institutional Review Board approval code 348-CS213.

[13]We included pop-up windows to verify that the participants were paying attention; the attention checker emerged after five minutes of inactivity in two randomly allocated screens. Once this pop-up window came up, the subject had to confirm she was still connected. If she did not validate her active status, her session concluded. We detailed this procedure in the general instructions and lost two participants for this reason (less than 2%).

included information on physician's profits and on that particular patient's benefits at every medical service level (see Figure A.2, in the Appendix). At this point, subjects had to decide on the number of medical services to allocate to the current patient. Once they had chosen, they could continue selecting the allocation for the group's second patient and then for the group's third patient.[14] This procedure is the same in both parts and for all treatments.

Once subjects had finished the decision rounds in a physician's role, which produced our variables of interest, we collected additional information on preferences for risk, prudence,[15] and altruism.[16] The activities for collecting this information also represented monetary payoffs for our subjects. Last, participants answered an unincentivized socioeconomic questionnaire.

The participants spent an average of 105 minutes in the activity and received an average payment of COP 37.000 (about $10). They received feedback on the roles assigned and the earnings at the end of the experiment. The feedback screen only appeared once all subjects in a session had concluded their participation. We included a final survey in Google forms where participants provided their bank account data for receiving their earnings from the activity. We followed the experimental procedures in Zhao et al. (2020).

1.3 Hypotheses

We now consider the hypotheses that we test in our results. We focus on whether budget constraints and uncertainty affect the allocation decisions of physicians and on how these decisions are made. An essential design feature is that the incentive structure favors over-provision when resources are abundant (see Table 1.1). Furthermore, we expect under-provision as a consequence of the constraint. Hence, we are particularly interested in evaluating the magnitudes of the misallocation of resources with and

[14] We included the possibility for subjects to revise the group screen before deciding each allocation.

[15] For this we use two certainty equivalence lotteries, one for risk preferences and one for prudence (Tarazona-Gomez, 2004) (see Figures A.6 and A.7, in the Appendix). In this case, we had a maximum of 5 ECUs for the risk task and of 4.75 ECUs for the prudence task.

[16] We elicit altruism through the Dictator Game. In our version of the game, participants decide how much money from an endowment of 5 ECUs to allocate to an anonymous and randomly matched participant. Payments depend on the match and the randomly assigned role (dictator or receiver).

without constraint. It can be expected that when resources are scarce, participants will devote more attention to how these resources are distributed (”scarcity captures our attention”), which will reflect in more efficient allocations (Mullainathan and Shafir, 2013).

Hypothesis 1. Misallocation of resources decreases when subjects are constrained with or without uncertainty.

The higher efficiency in the allocations under scarcity will also allow patients to be closer to their maximum achievable benefit, such that,

Hypothesis 2. Patients' health benefit is greater when physicians face a budget constraint than when they do not face resource restrictions.

On the other hand, we argue that uncertainty does not affect the incentive structure for physicians to provide medical services. As such, we should not observe changes in the benefits of patients nor on the efficiency of allocations whenever subjects have incomplete information on the composition of a group.

Hypothesis 3. Uncertainty does not affect allocation decisions compared to a setting with complete information.

Since our design also includes heterogeneous patients in several dimensions, we expect physicians to consider these characteristics for their allocation and distributional decisions. In particular, when resources are scarce, physicians, on average, should be likely to prioritize patients with a higher health improvement potential (think of COVID-19). However, it is not uncommon that physicians tend to patients in the worse initial status first, as with clinical triage, where time, facilities, and personnel are scarce. Last, patients with the lowest service needs are easy to prioritize even when resources are limited.

Hypothesis 4. Patients with a greater capacity to benefit from the allocated medical service units, those with below-median initial health status, and below-median optimal medical services needed see an improvement in their prioritization.

1.4 Results

This section explores the decisions of 154 subjects in both parts of the experiment and by treatment.[17] We have 1386 observations corresponding to 9 decisions made by the subjects in each part of the experiment. We lose some observations in the regression analysis due to the inability to classify some subjects regarding their risk preferences.[18] We first answer if budget constraints and uncertainty affect the allocation decisions of physicians. Then we move to explore the mechanisms.

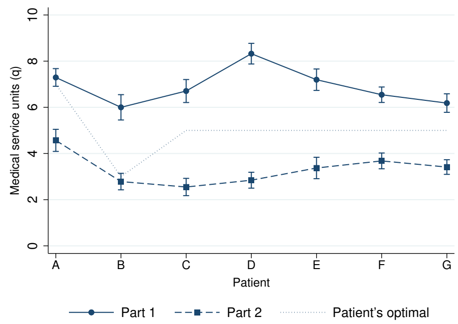
1.4.1 Do budget constraints and uncertainty affect allocation decisions by physicians?

We begin the analysis by plotting the average allocation to each profile, by part (Figure 1.2). Solid lines represent the allocations in Part 1, while the dashed line shows Part 2. The dotted line outlines the optimal level for each patient. From the figure, we can assess that, in the absence of constraints and uncertainty in Part 1, subjects allocate more medical service units to patients than what is optimal (over-provision). This result is consistent for all profiles ($p < 0.01$ using t-tests) and compatible with the experiment's incentive structure. In Table 1.3 we present further information to support this evidence. This result is line with other results in the literature (see Hennig-Schmidt et al. (2011), Brosig-Koch et al. (2016), and Brosig-Koch et al. (2017)).

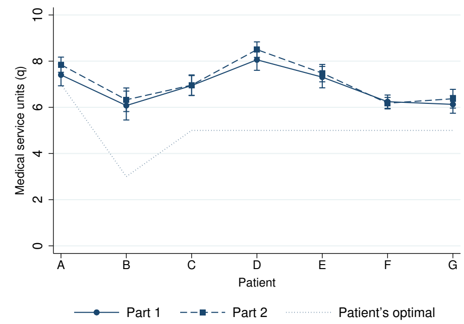
[17]Our groups are balanced in the socioeconomic characteristics within treatment (see C.6, in the Appendix). Moreover, we find no statistically significant differences at the 5% level in the distribution of the allocation decisions in our three treatments in Part 1 of the experiment ($p = 0.035$), using the Kruskal-Wallis test of equality of populations. Also, BC $\cup$ U is statistically different from BC+U ($p < 0.01$).

[18]Seventeen of our subjects (12.34% of our sample) have multiple switching points in our price list task for measuring risk-aversion. This rate of multiple switching is consistent with the levels found in the literature.

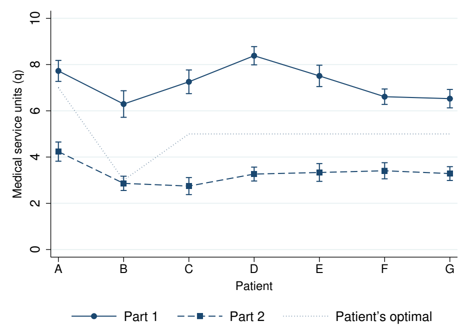
Figure 1.2: Medical services allocated by profile.



(a) BC



(b) U



(c) BC+U

Notes: These figures present the average quantity allocated to each patient profile by part of the experiment (y-axis). Panel (a) includes information for subjects in the Budget Constraint treatment; Panel (b) does the same for the Uncertainty treatment; and Panel (c) for Budget Constraint and Uncertainty. Solid lines relate to Part 1. Dashed lines correspond to Part 2. Dotted lines represent the $\hat{q}$ for each patient profile. Confidence intervals at the 95% using t-tests (x-axis).

Table 1.3: Medical services allocated, percentage deviation from patient's optimal level, and patient health benefits.

	BC			**U**			**BC+U**		
	Part 1	Part 2	$Part1 = Part2$	Part 1	Part 2	$Part1 = Part2$	Part 1	Part 2	$Part1 = Part2$
	(mean)		p-value	(mean)		p-value	(mean)		p-value
q	6.99	3.29	< 0.01	6.93	7.16	0.011	7.25	3.31	< 0.01
$\frac{q-\hat{q}}{\hat{q}}\%$	44.13	-32.98	< 0.01	42.98	47.60	0.018	49.43	-32.35	< 0.01
$B(q)$	8.26	8.58	0.236	8.41	8.34	0.412	8.07	8.64	0.013

Notes: Mean of units allocated, percentage deviation from units allocated to patients' optimal level, and patient's benefit by treatment and Part. BC stands for Budget Constraint, U for Uncertainty, and BC+U to Budget Constraint and Uncertainty treatments. P-values calculated from Wilcoxon matched-pairs signed-rank test.

Result 1: When information is complete and resources are abundant, there is over-provision to all patients' profiles.

The over-provision from Part 1 shifts to under-provision in Part 2 for subjects who face a budget constraint, but not to those facing only uncertainty (see Figure 1.1 and row 2 in Table 1.3). We perform a within-subjects analysis using a Wilcoxon matched-pairs sign-rank test. For BC and BC+U, we find that the number of units allocated to patients differs in both parts and is higher for Part 1 than for Part 2. This result is not surprising as the constraint forces downward subject decisions on how much they can allocate to patients.[19] Interestingly, in BC and BC+U, the average over-provision exhibited in part 1 is larger than the under-provision observed in part 2 ($p < 0.01$). For U, unlike expected, we find that the average medical service units assigned to patients is higher in Part 2 (at the 5% significance level). The same is true when we test the allocation from the patient's optimal level (see Table 1.3).

Table 1.4 presents additional evidence for these results. Here, we include OLS estimations for a within-subject analysis (difference in our outcome variables between parts). Columns (1), (3), and (5) include all patients for whom a physician makes an allocation decision. Columns (2), (4), and (6) exclude the third patients of each group. This exclusion guarantees that we only consider patients common to all physicians and treatments.[20] Controls include the patient's profile, subjects' age, gender, economics or finance major, socioeconomic status, if parents are medical professionals, and if they made mistakes in control questions. Our excluded category is U. We observe that considering each physician as her control, she allocates around 4 fewer units when she faces budget constraints, with or without uncertainty than those facing uncertainty alone. This result is almost mechanical. However, the relative allocation to the patient's optimum is reduced whenever physicians face a constraint, while it increases for those in U. Introducing constraints is generally reducing the inefficiency of allocations.

[19] 94.4% of medical decisions in BC and BC+U exactly exhaust the entire budget.

[20] We also tried specifications only excluding decisions to third patients in U and BC+U, and patients in BC when physicians had configurations 3 and 4 (see Table A.1, in the Appendix). We do not report these specifications as results are similar to those in columns (1), (3), and (5).

Table 1.4: Within-subjects analysis.

	$q_2 - q_1$		$\frac{q_2-\hat{q}}{\hat{q}}$ % - $\frac{q_1-\hat{q}}{\hat{q}}$ %		$B(q_2) - B(q_1)$	
	All	First two	All	First two	All	First two
	(1)	(2)	(3)	(4)	(5)	(6)
BC	-3.873***	-3.392***	-80.06***	-71.73***	0.332	0.395*
	(0.198)	(0.200)	(4.185)	(4.344)	(0.210)	(0.212)
BC+U	-4.226***	-3.931***	-87.64***	-83.29***	0.597**	0.769***
	(0.213)	(0.220)	(4.633)	(4.930)	(0.243)	(0.261)
Constant	1.158	1.413	37.89*	44.65**	-2.059*	-1.820
	(0.984)	(1.022)	(20.77)	(22.17)	(1.065)	(1.150)
Observations	1,215	810	1,215	810	1,215	810
Clusters	135	135	135	135	135	135
BC=BC+U (p-value)	0.142	¡0.05	0.139	¡0.05	0.291	0.168

Notes: OLS regressions. Dependent variables: within number of units allocated in columns (1) and (2), percentage deviation from patient's optimal level in columns (3) and (4), and patient's benefit in columns (5) and (6). BC: Budget Constraint; BC+U: Budget Constraint and Uncertainty; U: Uncertainty (reference category). Columns (1), (3), and (5) presents regressions for all patients and (2), (4), and (6) for the first two patients of each patient's group. Controls (all regressions): age, female, economics or finance students, socioeconomic status, if parents are medical professionals, mistakes in control questions, altruism, prudence, risk-aversion, and dummies for each patient profile. Standard errors clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Result 2: Budget restrictions, with or without uncertainty, reduce allocation to all patient profiles and cause under-provision. Notably, this under-provision is smaller (in absolute terms) than over-provision in the control condition and than under uncertainty.

Regarding the patient's health benefit (row 3 of Table 1.3), we observe that, on average, patients in BC+U receive allocations that lead to higher benefits in Part 2 than in Part 1 even when they receive fewer medical service units. Furthermore, in Table 1.4 we see that the first two patients in a group marginally receive higher benefits in Part 2 when treated by constrained physicians, instead of physicians in U. Interestingly, all patients receive higher benefits if physicians decide under uncertainty on top of constraints. The benefits are even higher when we only consider the first two patients in a group. A remarkable finding is that patient's benefits are higher for risk averse physicians in U. These results indicate that constraints generally lead to better conditions to patients when there are incentives to over-provide medical services.

Result 3: The benefits of patients increase between parts whenever physicians face constraints, with or without uncertainty, compared to those facing uncertainty only. On top of constraints, uncertainty favors all patients in the group in terms of relatively increasing their benefits.

1.4.2 How does this happen?

To study how uncertainty and constraints affect behavior, we perform a within-subject analysis that accounts for patient's characteristics and physician's prioritization criteria. We calculate a variable that ranks the position of each patient within a group according to the number of medical service units assigned, the percentage-wise absolute deviation from the optimal level,[21] and the patient's benefit.[22] The between-parts difference of these ranking variables reflect the redistributive effects of treatments in our outcomes variables, as they allow us to account for how the physician prioritizes among patients in a group. Then, our dependent variables are a discrete result of subtracting the rank in Part 1 from the rank in Part 2, leading to five ordinal categories. The lower category represents the case in which the patient at the top of the ranking in Part 1 drops to the bottom in Part 2. The higher category represents the opposite case.

Table 1.5 presents the results for the OLS regressions of the variables we described above for two different samples: decisions to all patients and decisions to the first two patients.[23] All regressions have our usual controls. The excluded category for the treatment variable is U. We include the four dimensions describing the health benefit functions of the seven patient profiles: initial health status (dummy of 1 when patients initial health benefit is above 5 UMEs, instead of less than 3), capacity to benefit from an additional medical service unit (dummy of 1 when the marginal benefit of

[21] We use the absolute deviation and not the percentage deviation we have been using up to this point. This is because we are interested in highlighting improvements in the distance to the patient's optimum, regardless of whether they are under or over-provided. For example, suppose that in the same group, profile A is under-provided by 3 units (-3), B is under-provided by 6 units (-6), and C is over-provided by 5 units (+5). Here, an ascendant ranking will assign a higher position to C than to B and to A than to B. However, A is closer to her optimum than both B and C. Using the absolute value, patient A achieves a better position than B and C in the ranking.

[22] Ties share the same position in the ranking.

[23] The results are equivalent when we treat these variables as categories. Estimations using ordered logit are available upon request.

an additional unit is 2 instead of 1), the number of medical services needed in the optimum (dummy of 1 when patients need 3 units of medical services for achieving their optimum, rather than 5 or 7), and the maximum benefit achievable (dummy of 1 when a patient's potential benefit at the optimum is above 10 UMEs, instead of 6). As can be seen, once we include patient's characteristics, those treated by physicians under constraint with or without uncertainty, improve their position in the ranking of allocations, but deteriorate in the deviation from the optimum compared to physicians in U.

Table 1.5: Patient's characteristics (within-subjects analysis)

	Rank q_2 - Rank q_1		Rank $\left\|\frac{q_2-\hat{q}}{\hat{q}}\right\|$ % - Rank $\left\|\frac{q_1-\hat{q}}{\hat{q}}\right\|$ %		Rank $B(q_2)$-Rank $B(q_1)$	
	All patients	First two patients	All patients	First two patients	All patients	First two patients
	(1)	(2)	(3)	(4)	(5)	(6)
BC	0.273*	0.815***	-0.371***	-0.388***	0.000779	-0.00515
	(0.157)	(0.162)	(0.135)	(0.127)	(0.119)	(0.0577)
BC+U	0.171	0.380**	-0.350**	-0.341**	0.0481	0.0930
	(0.176)	(0.168)	(0.166)	(0.154)	(0.115)	(0.0644)
Initial Health ≥ 5	0.0377	0.0444	0.00794	0	-0.187**	-0.0889
	(0.122)	(0.139)	(0.103)	(0.130)	(0.0805)	(0.0756)
Initial Health ≥ 5 × BC	-0.668***	-0.840***	0.798***	1.010***	0.502***	0.466***
	(0.205)	(0.222)	(0.174)	(0.200)	(0.124)	(0.125)
Initial Health ≥ 5 × BC+U	-0.719***	-0.788***	1.214***	1.439***	0.739***	0.711***
	(0.212)	(0.228)	(0.147)	(0.185)	(0.127)	(0.128)
Capacity to benefit of 2	0.0512	0.0444	-0.0968	-0.0889	-0.0128	-0.111
	(0.125)	(0.138)	(0.105)	(0.110)	(0.124)	(0.150)
Capacity to benefit of 2 × BC	0.355**	0.527**	0.331**	0.120	0.413**	0.448**
	(0.161)	(0.206)	(0.163)	(0.171)	(0.171)	(0.199)
Capacity to benefit of 2 × BC+U	0.0695	0.138	0.704***	0.479***	0.510***	0.538***
	(0.170)	(0.207)	(0.148)	(0.158)	(0.169)	(0.195)
$\hat{q}=3$	0.0295	0.0444	-0.0687	-0.0778	0.0690	-0.0778
	(0.134)	(0.169)	(0.103)	(0.144)	(0.0681)	(0.0771)
$\hat{q}=3$ × BC	0.735***	0.792***	1.035***	0.726***	1.281***	1.389***
	(0.205)	(0.267)	(0.200)	(0.243)	(0.146)	(0.154)
$\hat{q}=3$ × BC+U	0.708***	0.754**	1.324***	1.017***	1.305***	1.389***
	(0.246)	(0.327)	(0.198)	(0.256)	(0.135)	(0.131)
$B(\hat{q}) \geq 10$	-0.228*	-0.111	0.312***	0.189*	-0.168	0.122
	(0.134)	(0.164)	(0.102)	(0.114)	(0.112)	(0.0925)
$B(\hat{q}) \geq 10$ × BC	-0.132	-0.542*	-0.397**	-0.276	-0.636***	-0.709***
	(0.228)	(0.276)	(0.160)	(0.173)	(0.165)	(0.142)
$B(\hat{q}) \geq 10$ × BC+U	0.102	-0.0413	-0.799***	-0.689***	-0.873***	-0.958***
	(0.250)	(0.279)	(0.197)	(0.200)	(0.158)	(0.138)
Constant	0.120	0.251	-0.312	-0.199	0.350**	0.205
	(0.257)	(0.328)	(0.213)	(0.271)	(0.138)	(0.146)
Observations	1,215	810	1,215	810	1,215	810
Clusters	135	135	135	135	135	135
BC=BC+U (p-value)	0.580	0.0267	0.899	0.779	0.674	0.215

Notes: OLS regressions. Dependent variables: within group ranking of units allocated in columns (1) and (2), percentage deviation from patient's optimal level in columns (3) and (4), and patient's benefit in columns (5) and (6). In columns tagged "First two patients" we excluded the third patient from each group for the regressions. U: Uncertainty; BC+U: Budget Constraint and Uncertainty; BC: Budget Constraint. Reference categories: U, initial health status below 3, capacity to benefit of 1, maximum achievable benefit of 6, and $\hat{q} > 3$. Controls (all regressions): age, female, economics or finance students, socioeconomic status, if parents are medical professionals, mistakes in control questions, risk-aversion, prudence, altruism, and dummies for each patient profile. Standard errors clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

Overall, physicians under uncertainty alone do not appear to be treating patients differently according to their characteristics, except for some minor exceptions.

Result 4: Physicians do not respond to patients' characteristics when facing uncertainty alone.

However, constrained physicians, with or without uncertainty, do redistribute resources. Physicians reallocate resources towards patients with a relatively better initial health status, a higher capacity to benefit, and the lowest needs when moving from the control treatment to a situation with budget constraints, with or without uncertainty. In all cases, uncertainty, on top of constraints, represents more significant gains for these patients. These effects are not in place for physicians that move from the control to an uncertainty-only situation. The allocation ranking only worsens for those with an initial health benefit above 5 UMEs. Last, constrained physicians who face patients whose maximum achievable benefit is $B(\hat{q}) \leq 6$ (lowest possible among all profiles) prioritize them in terms of benefit and deviation to the optimum.

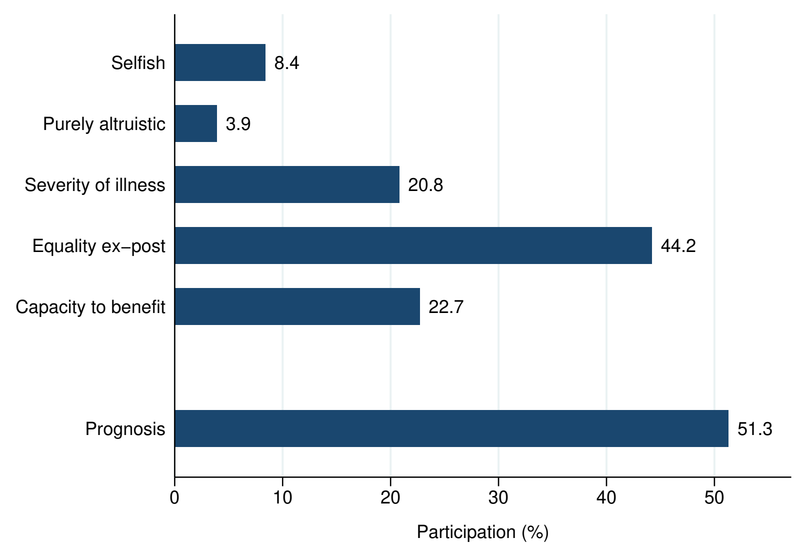
Result 5: Constrained physicians improve the prioritization of patients in relatively better conditions in terms of initial health status, capacity to benefit, and optimal needs, and those whose maximum benefit is the relatively lowest. If physicians also face uncertainty, on top of constraints, these patients are even better.

Next, we present the results of estimates where we include a classification of the physicians according to which patient they prioritize. This classification follows Martinsson and Persson (2019). We compare the allocation decisions to patients C and E in Part 1 and determine that physicians are selfish when they allocate $q = 10$ to both and purely altruistic if $q = 5$ for both C and E. Next, we classify subjects as prioritizing under equality ex-post when they allocate service units closer to the optimum to patient C compared to patient E. In contrast, subjects that favor the capacity to benefit allocate service units closer to the maximum benefit of E instead of C. Last, physicians whose prioritization criterion is the patient's severity of illness allocate the same number of units to both patients, but at a different level from selfish or purely altruistic physicians. In Figure 1.3, we present the frequency of these prioritization principles in our sample. Different to Martinsson and Persson (2019),[24] most of our

[24]Their sample is primarily classified as prioritizing under severity of illness and under capacity to

subjects (44% of the sample) follow an equality ex-post criterion when allocating with complete information and no constraints.

Figure 1.3: Distribution of physician types.



The figure presents the proportion of subjects (x-axis) classified in each principle of priority from Martinsson and Persson (2019) and prognosis (y-axis).

We formulate an additional principle referred to as prognosis (initial health status and maximum benefit achievable). When the allocation to profile D (E) is greater than the allocation to patient G (F), we categorize subjects as prioritizing according to the prognosis of patients. We classify physicians using this principle only for subjects whose behavior is consistent in D vs. G (capacity to benefit of 1) and E vs. F (capacity to benefit of 2). More than half of our subjects use the prognosis criterion, which means they over-provide more to patients with better prognoses.[25]

In Table 1.6 we present the OLS estimations for our variables of interest in Part

benefit.

[25]This criterion implies that physicians with incentives to treat are over-providing more to patients whose benefit will not suffer as much from that over-provision. When resources are ample, that means that prognosis will lead to a more conscious choice of how much to over-provide to patients whose prognosis is relatively worse.

2 vs. Part 1: allocation, deviation from the optimum, and patient's health benefits. Once more, we include the two patient samples we are considering (all patients and the first two patients in each group). All columns have the usual controls, and the excluded categories are U for the treatments and selfish for the priority setting principles.

From the results, it is clear that selfish physicians allocate fewer units between parts under constraint, with or without uncertainty than under uncertainty alone. Besides this seemingly mechanical result, the distance to the optimum also reduces in Part 2 concerning Part 1, and the patient's benefits increase. It is worth noting that changes in the between-parts allocations for all other physician types are not statistically different from those of selfish physicians in U. The same is true for the relative distance to the optimum and patient's benefits. These results mean that uncertainty alone does not appear to be affecting physician behavior, except for those classified as selfish. In addition, when we move to analyze the interactions between physician types and BC and BC+U, we confirm that patients tended by selfish constrained physicians see a reduction in the relative distance to the optimum and an increase in benefits, compared to all other types. This is a direct result of these patients being forced to make broader adjustments in their decisions regarding the control condition from Part 1.

Result 6: Selfish physicians behave closer to the patients' interests whenever they are resource-constrained, compared to when they are under uncertainty alone. If physicians are not selfish, treatments do not affect their behavior significantly.

Table 1.6: Physician's type (within-subjects analysis)

	$q_2 - q_1$		$\frac{q_2-\hat{q}}{\hat{q}}$ % - $\frac{q_1-\hat{q}}{\hat{q}}$ %		$B(q_2) - B(q_1)$	
	All	First two	All	First two	All	First two
	(1)	(2)	(3)	(4)	(5)	(6)
BC	-6.338***	-6.056***	-133.8***	-131.7***	3.286***	3.775***
	(0.392)	(0.363)	(8.160)	(7.724)	(0.593)	(0.465)
BC+U	-6.044***	-5.945***	-129.3***	-131.5***	2.869***	3.064***
	(0.451)	(0.464)	(9.589)	(9.620)	(0.595)	(0.509)
Priority concern:						
Purely altruistic	-0.188	-0.0831	-7.606	-7.419	-0.563	-1.106
	(0.536)	(0.599)	(11.44)	(13.22)	(0.699)	(0.721)
Purely altruistic × BC	4.491***	3.960***	96.21***	88.78***	-4.560***	-4.719***
	(0.553)	(0.685)	(11.87)	(14.99)	(0.784)	(0.791)
Purely altruistic × BC+U	4.096***	3.603***	93.64***	89.65***	-3.792***	-3.538***
	(0.649)	(0.763)	(13.79)	(16.20)	(0.829)	(0.852)
Severity of illness	0.231	0.139	2.839	0.113	-0.231	-0.389
	(0.438)	(0.391)	(9.146)	(8.286)	(0.516)	(0.453)
Severity of illness × BC	2.511***	2.576***	54.40***	57.78***	-2.837***	-3.260***
	(0.563)	(0.516)	(11.75)	(10.88)	(0.738)	(0.630)
Severity of illness × BC+U	1.358**	1.396**	30.99***	33.68***	-2.150***	-2.129***
	(0.547)	(0.562)	(11.53)	(11.48)	(0.724)	(0.694)
Equality ex-post	0.449	0.253	8.502	4.334	-0.209	-0.135
	(0.474)	(0.439)	(10.10)	(9.873)	(0.538)	(0.483)
Equality ex-post × BC	2.746***	2.875***	59.25***	64.00***	-3.315***	-3.911***
	(0.615)	(0.607)	(12.89)	(13.21)	(0.745)	(0.631)
Equality ex-post × BC+U	1.478**	1.532**	33.26**	36.18***	-1.878**	-1.832**
	(0.618)	(0.654)	(13.10)	(13.80)	(0.750)	(0.724)
Capacity to benefit	0.117	-0.0321	1.630	-1.716	-0.419	-0.609
	(0.452)	(0.414)	(9.660)	(9.137)	(0.540)	(0.507)
Capacity to benefit × BC	1.589**	2.266***	31.84**	45.43***	-1.964**	-2.567***
	(0.620)	(0.533)	(12.85)	(12.18)	(0.792)	(0.674)
Capacity to benefit × BC+U	1.274**	1.811***	27.85**	39.80***	-1.818***	-1.617***
	(0.554)	(0.609)	(11.70)	(12.51)	(0.653)	(0.604)
Overall prognosis	-0.368	-0.301	-8.273	-7.385	0.133	-0.00326
	(0.299)	(0.329)	(6.731)	(7.734)	(0.307)	(0.280)
Overall prognosis × BC	0.00876	0.0405	1.995	3.543	-0.110	0.0599
	(0.460)	(0.489)	(9.964)	(11.05)	(0.460)	(0.424)
Overall prognosis × BC+U	0.908**	0.968*	21.89**	24.95**	-0.853	-0.980*
	(0.443)	(0.489)	(9.442)	(10.59)	(0.524)	(0.513)
Constant	1.650*	1.969**	51.32***	61.15***	-2.591**	-2.227**
	(0.917)	(0.965)	(19.10)	(20.49)	(1.033)	(1.016)
Observations	1,215	810	1,215	810	1,215	810
Clusters	135	135	135	135	135	135
BC=BC+U (p-value)	0.345	0.784	0.478	0.975	0.446	0.0756

Notes: OLS regressions. Dependent variables: within number of units allocated in columns (1) and (2), percentage deviation from patient's optimal level in columns (3) and (4), and patient's benefit in columns (5) and (6). In columns tagged "All" we include all patients; in columns tagged "First two" we excluded the third patient from each group for the regressions. U: Uncertainty; BC+U: Budget Constraint and Uncertainty; BC: Budget Constraint. Reference categories: U and selfish, for physician type. Controls: age, female, economics or finance students, socioeconomic status, if parents are medical professionals, mistakes in control questions, risk-aversion, prudence, altruism, and dummies for each patient profile. Standard errors clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1.

1.5 Discussion and conclusions

We propose a decision-making experiment in which subjects decide as physicians on the sequential allocation of resources. We evaluate if and how constraints and uncertainty on the patients' relative needs affect behavior. We find that when physicians do not face any constraint or uncertainty, they allocate around 7.06 medical services to patients. Hennig-Schmidt et al. (2011), Brosig-Koch et al. (2016), and Brosig-Koch et al. (2017) found an overprovision equal to 6.6, 6.91, and 7.11 respectively, when they have similar incentives to ours. It appears that the design features we introduce, namely grouping patients by three and making the resources available for allocating to each group salient, do not affect average behavior in the lab. Once we include resource restrictions, we find that, with or without uncertainty, the over-provision previously observed becomes an under-provision for most patient profiles. In contrast, when physicians have incomplete information on the characteristics of the patients from the group but are not constrained, over-provision increases on average. Although some of these results can be considered somewhat mechanical, we find that the efficiency loss from the under-provision is smaller than for the over-provision and that patients can achieve relatively higher health benefits. Hence, we argue that when physicians have incentives to treat and are in the very probable scenario of not having complete information on the characteristics of the patients, there are gains in efficiency and health from the resource restrictions. We add to the existing literature, which centers on studying how financial incentives can help with a more efficient service provision (Hennig-Schmidt et al. (2011); Brosig-Koch et al. (2016, 2017), and others).

Furthermore, patients' characteristics interact with constraints and uncertainty, influencing physicians' behavior and forcing a within-group redistribution. In these situations, physicians reallocate in terms of efficiency and benefits towards patients in a relatively better situation. Patients whose benefit in the absence of medical services is relatively higher, those with twice the marginal benefit from an additional unit and who need the least number of services to be at their optimum, receive more efficient allocations and have higher benefits. The distributional decisions of physicians indicate that under these circumstances, resources and care are destined towards those who can achieve more with less and have the highest improving potential. Although constraints originate these results, the inclusion of uncertainty magnifies them. A final result comes from the physician's type. After determining physicians' priority setting principles under resource abundance and complete information, we observe that selfish physicians come closer to the patient's optimum and favor higher health benefits only when introducing constraints. In the future, we wish to explore how heterogeneous physicians in terms of their technology functions and other sources of uncertainty can change our findings.

Overall, budget restrictions in a medical decision-making setting improve patients' outcomes whenever physicians have incentives to over-treat. We do not intend to state that health systems should impose further constraints on physicians to secure better patient outcomes. Instead, we want to note that patients can indeed be better off in this situation than when resources are abundant, but we do not explore until what point this is true. This paper sets a restriction that guarantees that most patients can achieve some health benefit if physicians have distributive concerns. Future research could include considerations on how different constraint levels affect patients' well-being. In addition, physicians respond to patients' characteristics and reallocate towards patients who can achieve better outcomes with scarce resources. Interestingly, the better results intensify in the more realistic case of uncertainty and constraints. Last, the restrictions will only affect the decisions of selfish physicians. As a result, introducing restrictions on the number of services available when there are incentives to over-provide and when this over-provision is harmful to patients will generally improve patients' conditions and only affect physicians with no patient-regarding preferences.

Chapter 2

Social exclusion in the lab

2.1 Introduction

Bullying, broadly defined as a repeated exposition of injuries or discomfort from peers, is a primary threat to school safety and poses several undesirable long-term consequences. Aggression can happen either directly, in the form of words or physical contact, or indirectly, through ostracism and rumor spreading (Olweus, 1997). Bullying victims are more vulnerable to suffering from mental (depression, anxiety, and alcohol and drug abuse into adulthood) and physical (gastrointestinal distress and headaches) health issues (Lereya et al., 2015; Rivara and Le Menestrel, 2016; Olweus, 1997). Moreover, bullying increases school dropout and affects skill accumulation, and the perception of own exclusion correlates with lower achievement in math, language, and science (Berniell et al., 2016). Victimization also reduces non-cognitive skill accumulation and increases the probability of new bullying events (Sarzosa, 2021). Bullying victims in the lower decile of skill distribution are more likely to smoke, suffer from mental health diseases and depression, and have a lower chance of attending college (Sarzosa and Urzúa, 2021). Since human development and social mobility critically depend on health, cognitive, and non-cognitive skills (Heckman and Mosso, 2014), studying this phenomenon is of particular interest.

We propose a laboratory experiment to address a particular form of bullying in an experimental environment, namely social exclusion or ostracism. In particular, we want to study how social exclusion affects all the parties involved in the negative social interaction (offender, victim, and bystanders) regarding performance on a task and self-reported emotions. In our design, we depart from the study of how individual characteristics, including cognitive and non-cognitive skills, trigger ostracism (Olweus, 1997; Smith and Brain, 2000; Sarzosa and Urzúa, 2021) and focus on how exclusion in a social interaction setting generates changes in behavior. In that regard, we converse with the very sparse economic literature on the effects of bullying. In addition, we engage with the psychology literature on the effects of ostracism. We also explore the effect of bystander intervention through endorsement or punishment of the offenders' actions and the behavior of former victims once given a chance to become offenders.

We follow a series of experiments in psychology that use an ostracism manipulation (Cyberball) proposed by Williams et al. (2000). The original experiment invites participants to play a computerized ball-tossing game to improve their mental visualization skills, where the (usually two) other players are computer generated and controlled. In these experiments, the participants, who are unknowingly ostracized, report lower levels of belonging, self-esteem, control, and meaningful existence (Williams and Jarvis, 2006), even when a despised outgroup member excludes them (Gonsalkorale and Williams, 2007). These results seem to be more salient for adolescents, who, after facing exclusion from the game, report significantly higher anxiety and worse mood than adults in the same situation (Sebastian et al., 2010). Exclusion also activates brain regions associated with physical pain (Eisenberger et al., 2003; Cristofori et al., 2015). These effects are also present for bystanders (Polanin et al., 2012) whose neural pain network activates when seeing a social rejection of a friend during Cyberball (Beeney et al., 2011) (empathic response also in Masten et al. (2010)). In addition, Masten et al. (2011) find that participants with empathy-related brain activity behave more prosocially towards the victim after observing social exclusion. Riem et al. (2013) propose an adaptation of the Cyberball game to understand bystanders' behavior. In their case, the experimenter is ostracized by the computer, while the experimental subject is the bystander. Here, as well as in similar studies (Vrijhof et al. (2016)), they observe more throws towards the ostracized experimenter as a sign of prosocial compensating behavior.[1]

In our experiment, subjects connected to a virtual session participate in three parts consisting of individual and group stages. In all individual stages, subjects take part

[1] Regarding bystander participation, Gutierrez et al. (2018) evaluate the effectiveness of a government program aimed at increasing peer participation after bullying events.

in an encryption task (Benndorf et al., 2019)) that serves as our performance measure and answer two questions about the valence and arousal dimensions of their emotions. In group stages, four randomly matched participants interact in a modified version of the ostracism manipulation. We randomly assign them to a role denoted by a letter (A, B, C, or D) that determines the actions they can undertake during the group stages where the ball-tossing game happens, making them potential victims, offenders, or bystanders. In particular, A players (potential offenders) have monetary incentives to exclude B players (potential victims) during some group stages, and C and D players (bystanders) can sometimes rate player A's actions. In addition, we have group stages with role reversal for A and B.

We find that social exclusion events do not affect potential offenders' performance; however, the decision to exclude affects the valence dimension of emotions. In particular, the potential offenders that decide to include potential victims in all rounds report a reduction in valence compared to baseline, while those that decide to exclude at most twice report feeling relatively better. These changes in emotions are consistent with our design feature, where excluding other participants increases the potential offenders' monetary payoffs. On the other hand, we do not observe consistently significant changes in the performance of potential victims of the exclusion. There is an increase in the performance of non-excluded victims and a negative change concerning baseline for those excluded at most twice. The age of the participants explains the first effect; the arousal that appears after exclusion explains the second. Victimization does negatively affect the reported valence of emotions from ostracized participants, with the worst results associated with the highest level of social exclusion. There are no results for bystanders after social exclusion regarding performance or emotions. However, here are some hints that the two bystander types (C and D) are oppositely affected by ostracism, even though they are symmetric in their possibilities in the game.

Once we include the possibility for non-costly punishment, we find that even though the net effect to an offender that excludes a victim from the game and is harshly punished by the bystanders is zero, potential offenders change their behavior in response to the signals they get from the bystanders. Generalized disapproval of exclusion favors changes towards inclusion in the game, while generalized approval of inclusion maintains that behavior. In addition, punishment negatively affects the reported valence dimension of players' emotions in the offender, victim, and the bystander closest to the victim (C), and it positively affects the emotions of D bystanders. The effects on emotions do not translate to the performance in the task for most players. When punishment is below the median, the offenders and the participants in the victim position outperform themselves compared to the baseline. We argue that this lack of expected results may be explained by the income effect that follows from an incentivized task for measuring performance that can compensate for the monetary losses due to social exclusion.

Last, we find that subjects that were excluded when they were in the role of victim tend to exclude more intensely once they are in the role of offender. The emotions that were developed when they were victims of the exclusion and personality traits do not explain this behavior. Ostracism seems to be creating negative reciprocity towards others that are not at fault for the previous exclusion. We also observe that former potential victims that decide to exclude former potential offenders see a reduction in their performance compared to those that do not exclude. However, the results increase if they were excluded as victims. The opposite happens to former potential offenders. These results are in contrast with the absence of effects identified when we analyzed social exclusion behavior before role reversal.

The rest of the paper is organized as follows. In Section 2, we describe the experimental design and procedures. Then, in Section 3 we move to detail our hypotheses. In Section 4, we present the results. In Section 5, we discuss results and breakdown our conclusions.

2.2 Experimental Design and procedures

We propose an experimental design that involves three parts. All parts are composed of individual and group stages. Subjects participate in a Real Effort Task (RET) during the individual stages (six RET in total). This task, which gives us our performance measure, entails encoding three-letter words into numbers (Benndorf et al., 2019) (see Figure B.1 in the Appendix). In some of these individual stages, we also collect self-reported emotions using the Self-Assessment Manikin techniques from Bradley and Lang (1994) (see the diagrams in Figure B.2 in the Appendix). The emotional state includes valence and arousal measures, which are re-escalated from negative to positive for analysis purposes.[2] Group stages consist of a virtual tossing-the-ball game that follows the ostracism manipulation by Williams et al. (2000), and that lasts for five rounds. Contrary to Williams et al. (2000), participants in our Cyberball game are all connected to the virtual laboratory and interact with three other players also connected to the session. There are five group stages with particular rules that are common knowledge for all subjects right before the beginning of each. We also include a sample of questions from the Big Five Inventory (John et al., 1991) adapted for Colombia (Salgado et al., 2016) at the beginning of the experiment.[3] After Part 3, we include an incentivized questionnaire to elicit the social norm of exclusion (Krupka and Weber, 2013) and a sociodemographic survey.

Part 1 begins with an individual stage (RET_0) that we consider a trial stage, followed by the first group stage. In this first group stage, the computer randomly assigns participants to the letter A, B, C, or D. This letter determines the actions a participant can undertake in group stages and remains the same for all group stages in parts 1 and 2. During the first group stage in Part 1, we introduce the Cyberball game and the binding rule of tossing the ball only to the player on the right. This rule applies for the five rounds. A round starts when player A tosses the ball to player B and ends when player D tosses back the ball to player A (see Figure 2.1). We pay participants 5 Experimental Currency Units (ECUs) to toss the ball and 5 ECUs to receive the ball. This game dynamic yields a payment of 10 ECUs to each participant per round (50 ECUs during the group stage of Part 1). Once the five rounds of the first group stage end, we ask participants to report their emotions ($Emotions_1$) and to take part in a new individual stage (RET_1), which are our baseline measures for emotions and performance, respectively. In both individual stages of Part 1, participants earn 1 ECU for each correct encryption. The sum of the 50 ECUs from the group stage and the individual earnings made with the two RET from Part 1 make up for part of the gains from the activity.

[2]The nine scale figures are re-escalated from -4 to +4.

[3]We include two questions per dimension of the Big Five: for extraversion, we include "Is reserved" and "Is outgoing, sociable"; for agreeableness, we ask if "Is generally trusting" and "Tends to find fault with others"; for conscientiousness, we inquire if "Tends to be lazy" and "Does a thorough job"; for neuroticism, we ask if "Is relaxed, handles stress well" and "Gets nervous easily"; and for openness to experience we include "Has few artistic interests" and "Has an active imagination". With the answers to these questions we build an index using a Confirmatory Factor Analysis (CFA).

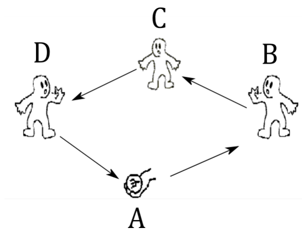


Figure 2.1: Group Stage from player A's point of view

Notes: The group stages consist of a virtual tossing-the-ball game that lasts for five rounds. A round begins when player A tosses the ball and ends after player D tosses the ball back to player A. Different group stages allow for different possibilities for the players.

Part 2 is composed of two group stages, each followed by an emotions questionnaire ($Emotions_2$ and $Emotions_3$, respectively) and performance tasks (RET_2 and RET_3, respectively) (see Figure 2.2). During the first group stage of Part 2, participants are part of the same group as in Part 1. Here, we introduce a monetary incentive for player A to exclude player B by tossing the ball directly to player C. If player A decides to follow the rule defined in Part 1, she earns the same 10 ECUs per round. However, if she decides to toss the ball directly to player C, she earns 15 instead of 5 ECUs for the toss and 5 ECUs for receiving the ball, for 20 ECUs per round. This game rule makes player A our potential offender in the game, player B the potential victim, and players C and D the (passive) bystanders. Note that if player A excludes player B from the game, player B cannot toss the ball to player C, and thus, she obtains 0 ECUs in that round. During the second group stage from Part 2, the group composition changes. In particular, all B, C, and D players remain part of their former group. However, the computer assigns player A to a new group. To facilitate this reassignment, we create super-groups of eight participants, among which group changes occur. We use these super-groups in all subsequent group stages for changes in group composition.

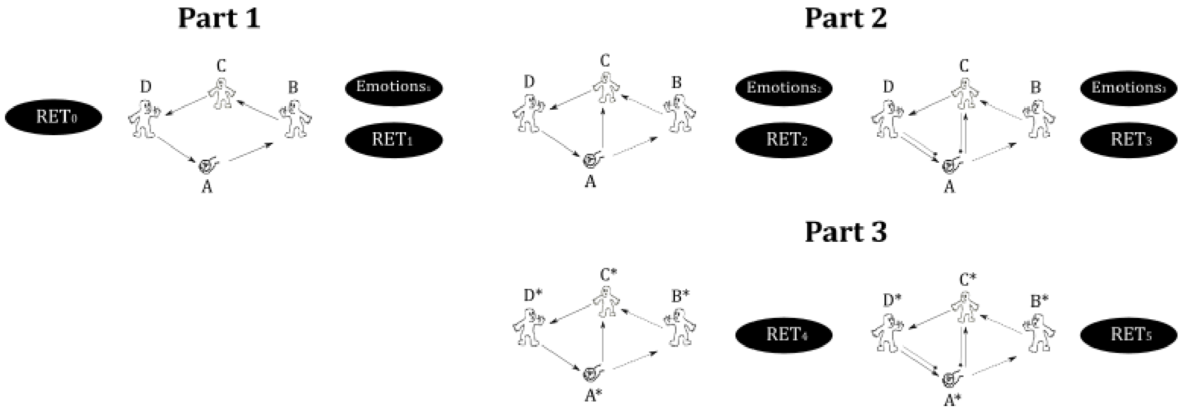


Figure 2.2: Experimental Design

Notes: Our experiment has three parts consisting of individual and group stages. The experiment begins with a RET that makes up for a trial stage, followed by a group stage that introduces the game dynamic of tossing the ball to the player on the right. After five rounds of a tossing-the-ball game, we collect our baseline measures for performance (RET_1) and emotions ($Emotions_1$). Part two has two group and individual stages. During the first individual stage from Part 2, we create monetary incentives for player A to deviate from the game rule and exclude participant B from the game. After 5 rounds, we collect measures for performance (RET_2) and emotions ($Emotions_2$). The second group stage introduces the possibility of punishment from C and D to A. Once five rounds are done, we collect new measures for performance (RET_3) and emotions ($Emotions_3$). Part 3 follows the same rules and dynamic as Part 2, but there is role reversal: former A players become B players, and vice versa. C and D players may or may not change roles among themselves. After the group stages in Part 3, we collect performance measures (RET_4 and RET_5, respectively).

Once we reassign A players to new groups, we again give player A incentives to exclude player B. During this stage, however, players C and D (the bystanders) must rate player A's actions using a 4-point Likert scale once she makes her decision. These ratings, which range from "Strongly Disagree" to "Strongly Agree," affect player A's earnings. For each "Strongly Disagree" or "Disagree" rating, player A faces a reduction of 5 and 3 ECUs in her payments, respectively. Analogously, for each "Strongly Agree" or "Agree" rating, player A's earnings increase by 5 and 3 ECUs, respectively. If players C or D rate player A with "Neither Agree nor Disagree," there are no changes in player A's earnings. The ratings become common knowledge to all players before starting the next round. Note that the distance to the social exclusion is different for C and D bystanders. In particular, the C bystander is closest to the potential victim and is the one that receives the ball when player A excludes her from the game. Oppositely, D bystander's game dynamic does not change much with or without social exclusion. These differences can have implications on how much they are affected by victimization.

Part 3 follows the same rules and structure as Part 2. However, there are changes in the roles previously assigned to participants. In particular, former A players become B players, while former B players become A players; we randomly resort C and D players into D or C players. With this manipulation, we can study how former victims of the exclusion behave when they can ostracize a group member and how former offenders react to this exclusion. We use the same super-groups of eight participants defined earlier and rematch them into two new groups once the roles are reassigned. This means that the group configuration changes between parts 2 and 3. In addition, Part 3 does not include self-assessment techniques to elicit changes in emotions. The computer randomly selects Part 2 or Part 3 to determine the final earnings from the activity. The earnings during the individual and group stages from the part selected by the computer are added to the earnings from Part 1 to determine final payoffs.

We performed 13 virtual sessions with students from Universidad del Rosario. Each session consisted of 24[4] participants connected to a Zoom session. We invited participants to an activity that would take up to two hours, including an identity check-up. The experiments were programmed in oTree (Chen et al., 2016), and we recruited subjects using the Online Recruitment System for Economic Experiments (ORSEE) from Rosario Experimental and Behavioral Economics Lab (REBEL). Once in the session, we admitted participants from a waiting room in groups of four and modified their screen names with a code to maintain anonymity. We randomly selected five participants from each session for identity check-ups using Zoom's private rooms. Once the experiment started, We read aloud the General Instructions and the instructions for the first individual and group stages. The rest of the instructions were available on the experimental screens. Two experimenters were available for questions through the Zoom chat. We included inactivity alerts in all rounds from the group stages. We lost three participants due to inactivity and dropped all the other members of their respective supra-groups.

We held a first group of ten sessions between October 14 and November 19, 2021. We collected data in three more sessions on May 2nd and 3rd, 2022. This second group of sessions was necessary to replace the observations we had to drop due to inactivity

[4]Session 12 was carried out with 16 participants.

(see the Results section for details on the final sample). The participants spent 89 minutes in the activity and received an average payment of COP 60.000 (about $15). Once all participants concluded their participation, they received feedback on the Part selected by the computer for payments and their earnings. At the end of the feedback screen, we included a link for a form where we collected information on the bank account where subjects wanted to receive their earnings from the activity.

2.3 Hypotheses

Our hypotheses follow some of the main findings of the literature on bullying. We study as our primary outcomes a measure that follows from the performance on a RET and self-reported measures of emotions (valence and arousal). We first center on the effects of social exclusion on direct victims and bystanders. We understand exclusion in its intensive margin, meaning the number of ostracism events and some defined categories for the intensity of the exclusion. Following the findings, we expect that:

Hypothesis 1. Social exclusion events will negatively affect the performance in the RET and the self-reported emotions of both victims and bystanders. They will not affect the performance and emotions of offenders.

Although the literature on the effects of bullying on bystanders indicates the existence of pro-social compensating behavior, the closeness between victims and bystanders and the individual prosociality levels of bystanders mediate these results. Then,

Hypothesis 2. The effects of social exclusion in performance and emotions are greater for direct victims of the exclusion than for bystanders.

Our experimental design includes the possibility of punishment through ratings. These ratings can potentially affect the earnings of player A in at most 10 ECUs per round if both C and D players decide on the maximum punishment for player A. A payoff maximizing offender should be indifferent between tossing the ball to player B or player C when she believes the bystanders will harshly punish her excluding decision. In fact, she can increase her expected earnings by excluding player C. However, the presence of ratings can, by itself, discourage exclusion (Fehr and Gächter, 2000):

Hypothesis 3. When both bystanders punish the offender, this reduces exclusion in upcoming rounds.

In addition, the degree of punishment should affect the RET performance and the reported emotions of the directly implicated parties (offender and victims). However, since there are no costs of the punishment to the bystanders, we should not observe changes in the outcomes of interest.

Hypothesis 4. Punishment negatively affects the RET performance and the reported emotions of offenders and positively affects those of victims. These results remain stable for the bystanders.

Once there is a role reversal in Part 3, we expect that:

Hypothesis 5. Victims who were excluded more frequently in Part 2 and who have the offender's role in Part 3 will exclude their victims more often.

And,

Hypothesis 6. The offenders who are in the victim role in Part 3 will have minor variations in the RET as long as they were part of the top excluders. However, if they excluded below the median and face repeated exclusion, they will have worse results in the RET.

In addition, we want to explore what happens with the RET performance of victims in Part 3, depending on their former behavior as bullies in Part 2.

2.4 Results

We study the decisions of 278 subjects[5] in five parts. The computer initially assigns subjects to roles A (N=69), B (N=70), C (N=69), and D (N=70) and reassigns them

[5] 304 subjects participated in twelve sessions of 24 participants and one of 16. We eliminate 26 subjects: five dropped participants due to inattention, three research assistants that replaced participants who left the session before starting in order to be able to begin on time, and the other members of their super-groups. A dropped participant is replaced by the computer and decisions are determined at random.

to new roles from part 4 onward, as detailed earlier.[6] We begin our analysis with Table C.6, which presents the summary statistics by the subjects' initial roles. We include our main sociodemographic variables, relevant academic information, in index factoring the Big Five test dimensions, the mean ratings for the elicited social norms,[7] the baseline measures from the emotions test (valence and arousal), and the baseline and trial RET performance. Overall, there is balance in these variables across initial roles, as shown by the p-values from the mean comparison tests. There is also no evidence of a difference in the baseline RET or the emotions test by role pairs according to the Mann-Whitney test, which is confirmed with the Kruskal-Wallis test ($\chi^2 = 3.084$ and $p-value = 0.358$ for the baseline performance in the RET, and $\chi^2 = 1.709$ and $p - value = 0.635$ for the baseline reported emotions).

[6] A (N=70), B (N=69), C (N=69), and D (N=70).

[7] To convert the likert scale responses into numerical scores we follow Krupka and Weber (2013) and define that: "very socially inappropriate" = –1, "somewhat socially inappropriate" = –1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1.

Table 2.1: Summary statistics and mean comparison tests, by role

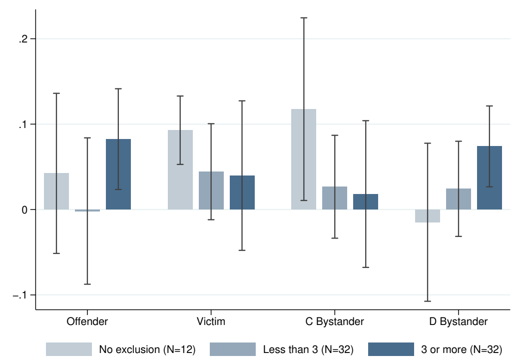
	Mean				p-value					
	A (N=69)	B (N=70)	C (N=69)	D (N=70)	A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
Socio-Demographic Characteristics:										
Age	21.00	21.27	21.01	21.60	0.476	0.963	0.132	0.493	0.463	0.136
Female	0.55	0.61	0.67	0.63	0.450	0.164	0.354	0.523	0.863	0.641
Number of siblings	1.54	1.30	1.38	1.56	0.173	0.413	0.913	0.673	0.148	0.364
Number of household members	3.83	3.60	3.43	3.36	0.288	0.118	0.033	0.494	0.246	0.753
Relative income at 16	3.32	3.49	3.54	3.41	0.232	0.102	0.451	0.705	0.576	0.309
Above-average income at 16	0.43	0.43	0.46	0.34	0.942	0.734	0.269	0.679	0.300	0.147
Strata	3.61	3.74	3.75	3.59	0.431	0.414	0.892	0.954	0.380	0.366
Self-financed expenditure	31.01	32.57	34.93	34.57	0.779	0.472	0.533	0.660	0.722	0.948
University information:										
Semester	6.91	6.59	6.59	7.07	0.322	0.329	0.615	0.979	0.114	0.115
Economics	0.20	0.10	0.12	0.16	0.091	0.164	0.486	0.764	0.315	0.482
Economics or finance	0.30	0.24	0.28	0.37	0.419	0.710	0.406	0.664	0.099	0.228
Big Five:										
Personality index	0.03	0.04	-0.05	-0.03	0.871	0.152	0.332	0.095	0.238	0.645
Social norms										
No exclusion	0.86	0.84	0.83	0.81	0.765	0.658	0.486	0.869	0.694	0.835
One round of exclusion	0.50	0.45	0.36	0.31	0.485	0.054	0.015	0.274	0.107	0.555
Two rounds of exclusion	0.22	0.20	0.16	0.10	0.809	0.432	0.132	0.563	0.178	0.418
Three rounds of exclusion	-0.29	-0.23	-0.26	-0.32	0.458	0.685	0.609	0.710	0.224	0.358
Four rounds of exclusion	-0.67	-0.60	-0.64	-0.62	0.340	0.706	0.478	0.594	0.804	0.766
Five rounds of exclusion	-0.90	-0.96	-0.90	-0.90	0.240	1.000	0.891	0.221	0.213	0.888
Baseline measures										
Trial Real Effort Task result	7.20	7.59	7.70	6.91	0.254	0.097	0.414	0.747	0.087	0.030
Trial Real Effort Task mistakes	0.29	0.14	0.28	0.13	0.060	0.875	0.037	0.110	0.826	0.074
Baseline Real Effort Task result	8.29	8.70	8.54	8.36	0.121	0.381	0.787	0.595	0.221	0.544
Baseline Real Effort Task mistakes	0.38	0.29	0.23	0.40	0.474	0.251	0.871	0.628	0.379	0.192
Reported emotion	1.99	1.79	1.94	1.71	0.469	0.874	0.320	0.571	0.795	0.404
Reported intensity of emotion	0.32	0.53	0.32	0.30	0.503	1.000	0.948	0.552	0.453	0.954

Notes: This table shows the mean by the initial role assigned (A, B, C, and D) for the main final survey variables, the Big Five dimensions, and the baseline measures. It also includes the p-value for the mean differences by role pair. We present results for age, female proportion, number of siblings and household members, relative income at 16, above-average income at 16, and percentage of self-financed expenditures. We also add socioeconomic status following the Colombian classification. We include the current semester and if the participant is an Economics or Finance major. We build an index factoring the five dimensions of the Big Five personality test from a sample of questions. We include the average responses to the elicited norms after converting responses into numerical scores (we follow Krupka and Weber (2013) and define "very socially inappropriate" = –1, "somewhat socially inappropriate" = –1/3, "somewhat socially appropriate" = 1/3, "very socially appropriate" = 1). In addition, we present the baseline results from the encryption task and the number of mistakes, as the reported emotions and intensity of emotions at the beginning of the experiment.

2.4.1 Social Exclusion

We begin studying the effects of social exclusion events in the performance and emotions of all participants. During the first group stage from Part 2, we create monetary incentives for potential offenders (A players) to exclude B players (potential victims) from the ball-tossing game. From 345 rounds, corresponding to 5 rounds for each of the 69 potential offenders, we find exclusion in 49% of the opportunities. Once we include the possibility of punishment, exclusion reduces to 29.3%. Role reversal does not significantly change this dynamic as exclusion with no punishment is 46.3% and 32% with punishment. Figure 2.3 shows the average changes in the RET performance between Part 1 (RET_1) and the first group stage from Part 2 (RET), normalized by the performance during the baseline RET (RET_1).[8] To account for the intensity of the exclusion, we create three exclusion categories: no exclusion, exclusion in less than three rounds, and exclusion in three or more rounds. This categorization allows us to find the effects of below the median and above the median exclusion. In addition, it is consistent with the social norm elicitation results that show that a positive social valuation of exclusion corresponds to, at most, two rounds of exclusion. From three rounds of exclusion onward, there is an average agreement of social inappropriateness. We present results for the potential offenders and victims and both bystanders.

Figure 2.3: Average changes in normalized RET performance by exclusion category, by role.



Notes: This figure presents the average changes in the normalized RET performance of participants between Part 1 and the first group stage from Part 2. We differentiate results by role for each defined category for exclusion intensity (no exclusion, less than two, and three or more rounds). The number of observations per category available inside the parentheses. As there is only a possibility for exclusion, these are potential offenders or victims.Confidence intervals at the 95% using t-tests.

As can be seen, the normalized changes in performance are not significantly different from zero in most cases. Notably, non-excluded subjects in the potential victim role appear to outperform themselves compared to the baseline (Wilcoxon signed ranks test results are $z = 2.512$ and $p < 0.05$). The results are also significantly better for potential offenders ($z = 3.170$ and $p < 0.01$), potential victims ($z = 2.400$ and $p < 0.05$),

[8]With this normalization, we try to account for the individual ability in the task.

and D bystanders ($z = 2.949$ and $p < 0.01$) whenever heavy exclusion occurs (three or more exclusion events during the first group stage from Part 2). However, we do not find significant differences for the social exclusion categories within roles using the Kuskal-Wallis tests. Table 2.2 presents the OLS results for the normalized changes in performance for potential offenders and potential victims. Columns (1), (2), (5), and (6) consider the number of exclusions by part as the independent variable; the rest use the categorical variable previously defined for the exclusion, with no exclusion as the omitted category. Even columns include controls for changes in emotions, sociodemographic variables (age, gender, number of siblings, number of household members, and Colombian classification for socioeconomic status), an index of personality traits from the Big Five Inventory, and if the subject is an economics or finance major. We also include the number of times the computer had to toss the ball during the first group stage from Part 2 to measure inattention.[9]

Table 2.2: OLS results for within-subject changes in normalized RET performance by exclusion, for potential offenders and victims

	$\frac{RET_2-RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender				*Potential Victim*			
Number of exclusion events by part	0.006	0.003			-0.001	0.011		
	(0.011)	(0.016)			(0.012)	(0.015)		
Less than 3 exclusions by part			-0.017	-0.022			-0.063*	-0.036
			(0.062)	(0.078)			(0.035)	(0.043)
3 or more exclusions by part			0.041	0.031			-0.023	0.050
			(0.062)	(0.081)			(0.035)	(0.063)
Constant	0.041	0.307	0.044	0.286	0.062**	-0.190	0.096***	-0.157
	(0.038)	(0.375)	(0.055)	(0.362)	(0.028)	(0.221)	(0.021)	(0.194)
Observations	69	69	69	69	70	70	70	70
Clusters	69	69	69	69	70	70	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions for potential offenders and victims. Dependent variable is the change in the normalized RET performance in the first group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference category in columns (3), (4), (7), and (8) is no exclusion. Controls: changes in emotions, sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), an index for personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

These results align with what we expected for the potential offender. In particu-

[9] A detailed table for the effects of the control variables is available in the Appendix (Table B.1).

lar, exclusion events do not appear to affect the performance of the offenders in the RET. Unlike expected, the normalized RET performance of potential victims is also not affected by the number of rounds when the offenders excluded them. However, when we consider exclusion categories, we find that the RET result for potential victims significantly increases whenever they are not excluded. This result disappears once we control for sociodemographic characteristics, and the direction of the effect changes with the inclusion of age into the regression. A non-materialized possibility of exclusion is only positively affecting older participants. Excluded participants in the victim position observe a lesser improvement in performance than non-excluded B players; this finding is not statistically significant for those excluded in three or more rounds. Once we control for the changes in the reported emotions, this result stops being significant for those excluded at most twice and changes direction for the heavily excluded participants. However, it is still not statistically significant (see Table B.1 in the Appendix for details). The valence dimension of exclusion appears to be essential for explaining changes in the results of victims of the exclusion.

Result 1: Social exclusion events generally do not affect potential offenders in our task.

Result 2: When there are no exclusion events, potential victims outperform themselves compared to their baseline performance measure. This effect appears to be driven by older participants. In addition, lightly excluded victims see a reduction in their performance between parts compared to the non-excluded potential victims. The valence dimension of emotions explains this last effect.

Table 2.3 presents analogous results for the pooled bystanders. Columns (1) and (2) use the number of exclusions as the independent variable; columns (3) and (4) use the categories for exclusion previously described, with no exclusion as the reference category. We observe that when the number of exclusion events increases, the performance of D bystanders also increases. The opposite is true for C bystanders that see a reduction in their RET performance when the number of exclusions to B increases, although this result is not statistically significant. Similarly, when we advance in the exclusion categories, the performance of D bystanders increases, while it decreases for C bystanders. These results may hint that the relationship between the two bystanders and the victims is different. However, we cannot confirm statistically that the two bystanders behave differently when facing social exclusion or that, as hypothesized, their performance is negatively affected by victimization.

Table 2.3: OLS results for within-subject changes in normalized RET performance by exclusion, pooled bystanders

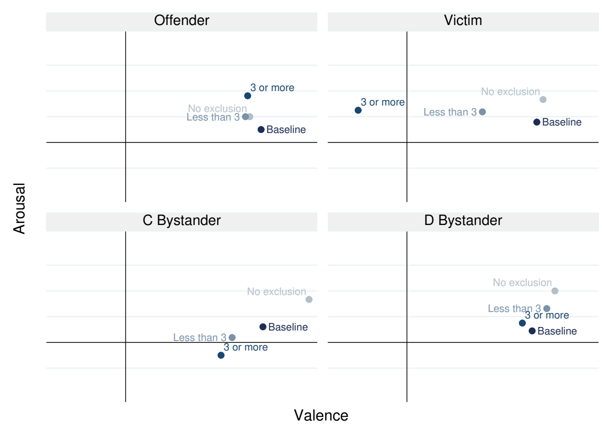
	$\frac{RET_2-RET_1}{RET_1}$			
	(1)	(2)	(3)	(4)
Number of exclusion events by part	0.019*	0.020*		
	(0.011)	(0.011)		
C Bystander	0.040	0.030	0.073	0.062
	(0.057)	(0.053)	(0.079)	(0.084)
Number of exclusions × C Bystander	-0.024	-0.021		
	(0.018)	(0.019)		
Less than 3 exclusions by part			0.029	0.027
			(0.060)	(0.063)
3 or more exclusions by part			0.080	0.085
			(0.058)	(0.061)
Less than 3 exclusions × C Bystander			-0.071	-0.063
			(0.094)	(0.100)
3 or more exclusions × C Bystander			-0.139	-0.128
			(0.094)	(0.108)
Constant	-0.002	-0.029	-0.004	-0.047
	(0.033)	(0.238)	(0.053)	(0.245)
Observations	138	138	138	138
Clusters	70	70	70	70
Controls	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the normalized RET performance in the first group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference category in columns (3) and (4) is no exclusion. Controls: changes in emotions, sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), an index for personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Result 3: The two bystanders appear to have contrasting performance effects when observing social exclusion, but the effects are not statistically significant.

We observe results consistent with our hypotheses in terms of the reported emotions. Figure 2.4 includes the two dimensions of the reported emotions: valence and arousal. We graph the average of the re-escalated measures from both dimensions that range from -4 to +4. This figure contains the baseline reported emotions and the results after the first group stage from Part 2. Overall, all subjects report positive high arousal emotions in the baseline. After the exclusion possibility, a Kruskal-Wallis test of equality of populations indicates that we can only reject the null hypothesis for the victims and only in terms of the valence dimension of emotions ($p < 0.01$). In particular, the changes in valence for those heavily excluded concerning baseline are statistically different from the changes for those not excluded (Mann-Whitney test results are $z = 3.897$ and $p < 0.01$) and from the changes in emotions for those excluded once or twice ($z = 3.083$ and $p < 0.01$). The results are also different for lightly excluded participants from those not excluded ($z = 2.251$ and $p < 0.05$). In addition, A players who do not exclude B participants report feeling worse than at baseline.[10] C bystanders who observe exclusion in three or more rounds also appear to have negative changes in the valence dimension of emotions. However, the same is not observable for D bystanders. The arousal levels also appear to increase with the levels of exclusion for players A and B.[11] However, for the C bystander, the arousal levels whenever there is exclusion reduce.[12]

Figure 2.4: Two dimensions of the reported emotions by intensity of the exclusion, by role



Notes: These figures present the average of the two reported dimensions in the emotion test: valence and arousal. We graph both dimensions by role and by group stage for each of the defined categories for exclusion intensity (no exclusion, less than two, and three or more rounds). The baseline corresponds to Part 1 of the experiment. We also include the results after the first group stage in Part 2 (Exclusion). As there is only a possibility for exclusion, these are potential offenders or victims.

[10]This is statistically different from the changes in valence reported by those who exclude once or twice during the first group stage from Part 2 ($z = -1.803$ and $p = 0.0714$).

[11]A within-subject comparison gives significant results for A participants who exclude three or more times (Wilcoxon signed-rank test gives $z = 4.179$ and $p < 0.01$), and for B participants heavily excluded (Wilcoxon signed-rank test gives $z = 4.211$ and $p < 0.01$).

[12]Results for the Wilcoxon signed-rank test gives $z = -2.242$ and $p < 0.05$ for the cases when there is exclusion once or twice and $z = -2.684$ and $p < 0.01$ for the cases where there are three or more exclusions.

Table 2.4 presents the OLS results for the two dimensions of the reported emotions by role. Columns (1) to (4) use the within-subject difference in the reported valence as the dependent variable; columns (5) to (8) report the results for the within-subject difference in the reported arousal. In columns (1), (2), (5), and (6), we use the number of exclusions as the independent variable; the rest of the columns use the exclusion categories previously defined. Odd columns include controls for sociodemographic variables, personality traits, economics or finance major, and a measure for inattention (detailed results including controls available in tables B.2 and B.3 in the Appendix). The standard errors are clustered at the group level. We first observe that, as mentioned earlier, subjects in the offender's role that do not exclude B players report a reduction in the valence dimension of emotions. This effect disappears once we control for sociodemographic variables and changes direction when we include controls for age. However, we find consistent evidence that offenders who decide to exclude report higher valence than their baseline measures. This result is statistically significant and higher in magnitude for those who exclude only once or twice during the first group stage from Part 2.

Table 2.4: OLS results for within-subject changes in reported emotions, for potential offenders and victims

	$Valence_2 - Valence_1$				$Arousal_2 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Potential Offender								
Number of exclusion events by part	-0.062	-0.055			0.151	0.149		
	(0.117)	(0.115)			(0.135)	(0.119)		
Less than 3 exclusions by part			0.693**	0.780**			-0.307	-0.350
			(0.338)	(0.312)			(0.481)	(0.557)
3 or more exclusions by part			0.233	0.335			0.136	0.041
			(0.413)	(0.377)			(0.494)	(0.526)
Constant	-0.006	3.400*	-0.556*	2.740	0.007	3.540	0.444	4.265
	(0.252)	(1.936)	(0.283)	(1.908)	(0.339)	(3.176)	(0.397)	(3.547)
Observations	69	69	69	69	69	69	69	69
Clusters	69	69	69	69	69	69	69	69
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Potential Victim								
Number of exclusion events by part	-0.727***	-0.665***			0.101	0.060		
	(0.173)	(0.187)			(0.117)	(0.141)		
Less than 3 exclusions by part			-1.178***	-1.189***			-0.144	-0.263
			(0.407)	(0.378)			(0.369)	(0.418)
3 or more exclusions by part			-3.057***	-2.933***			0.308	-0.035
			(0.523)	(0.574)			(0.461)	(0.589)
Constant	0.351	2.881	0.444	3.215	-0.059	-1.783	0.111	-1.572
	(0.345)	(2.488)	(0.283)	(2.520)	(0.280)	(1.925)	(0.251)	(1.933)
Observations	70	70	70	70	70	70	70	70
Clusters	70	70	70	70	70	70	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variables: changes in the reported valence and arousal in the first group stage from Part 2 compared to the baseline. Reference category even columns is no exclusion. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Result 4: Contrary to expected, exclusion affects the reported emotions of A players. In particular, exclusion increases our valence measure.

On the other hand, and as expected, victims report feeling worse than baseline as exclusions increases. The negative changes in valence are deepened as we move up in the exclusion categories. Table 2.5 presents the results for C and D bystanders. Here we see that C bystanders are negatively affected in terms of valence when there is social exclusion. The same is not true for D bystanders. There are no significant changes in the arousal dimension of emotions for any of the players. Tables B.2 and B.3 in the appendix present results detailing controls.

Table 2.5: OLS results for within-subject changes in reported emotions, for bystanders

	$Valence_2 - Valence_1$				$Arousal_2 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C Bystander								
Number of exclusion events by part	-0.245	-0.235**			-0.228	-0.208		
	(0.149)	(0.114)			(0.150)	(0.130)		
Less than 3 exclusions by part			-0.069	-0.546			-0.617	-0.655
			(0.441)	(0.409)			(0.530)	(0.587)
3 or more exclusions by part			-0.613	-0.720*			-0.993	-0.787
			(0.530)	(0.429)			(0.653)	(0.662)
Constant	0.291	-0.751	-0.000	-0.769	0.295	1.952	0.444	1.971
	(0.313)	(1.805)	(0.394)	(1.825)	(0.317)	(2.473)	(0.511)	(2.486)
Observations	69	69	69	69	69	69	69	69
Clusters	69	69	69	69	69	69	69	69
Controls	No	Yes	No	Yes	No	Yes	No	Yes
D Bystander								
Number of exclusion events by part	-0.062	-0.023			-0.121	-0.124		
	(0.107)	(0.105)			(0.099)	(0.099)		
Less than 3 exclusions by part			-0.767*	-0.647			0.278	0.265
			(0.452)	(0.486)			(0.368)	(0.416)
3 or more exclusions by part			-0.602	-0.449			0.036	0.038
			(0.523)	(0.535)			(0.340)	(0.367)
Constant	0.222	0.827	0.667	1.091	0.651**	1.268	0.222	0.833
	(0.256)	(1.882)	(0.425)	(1.855)	(0.319)	(1.955)	(0.142)	(1.960)
Observations	70	70	70	70	70	70	70	70
Clusters	70	70	70	70	70	70	70	70
Controls	No	Yes	No	Yes	No	Yes	No	Yes

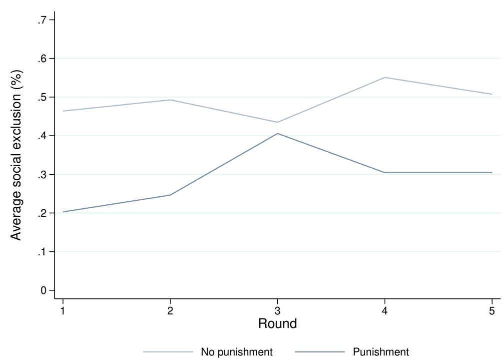
Notes: OLS regressions. Dependent variables: changes in the reported valence and arousal in the first group stage from Part 2 compared to the baseline. Reference category even columns is no exclusion. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Result 5: Exclusion negatively affects the reported valence of excluded players and of C bystanders, but not of D bystanders.

2.4.2 Punishment

We now move to analyze the punishment behavior of our subjects. On average, our offenders exclude in 2.45 of the five rounds whenever there is a possibility of exclusion without punishment. Once we include the possibility of punishment, the average exclusion reduces to 1.46 rounds. Using a Kuskal-Wallis test we reject the null hypothesis of equality of population between the exclusion behavior of potential offenders under the punishment and the no punishment conditions ($p < 0.01$). [13] Figure 2.5 shows the average social exclusion behavior of A players by round. As can be seen, ostracism is mostly stable in the no-punishment condition, varying between 43.5% and 55.1%. In the punishment condition, social exclusion drops to 20% in round 1, compared to 46% in the no-punishment condition. This may indicate that the threat of punishment disciplines potential offenders. The exclusion then rises to around 25% in round 2 and over 40% in round 3. It then stabilizes at 30%. This variability is most likely responding to the punishment decisions of the bystanders.

Figure 2.5: Average social exclusion with and without punishment, by round

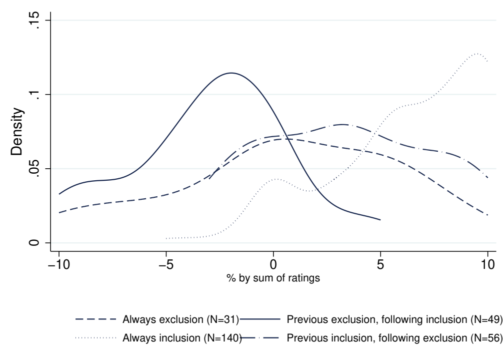


In an initial exploration of bystander behavior when there is a possibility of punishment we find that when there is social exclusion in a round, the sum of the opinions of both bystanders is negative and significant. When controls are included, we find that older subjects and those of a higher socioeconomic status give higher ratings to potential offenders (see Table B.4 in the Appendix). Figure 2.6 includes the density for the social exclusion decisions of potential offenders, depending on the sum of ratings from bystanders in the immediately previous round. The continuous line illustrates

[13] There is a significant difference in the exclusion behavior with and without punishment ($z = 4.128$ and $p < 0.01$) using a Wilcoxon signed-rank test.

that changes in social exclusion behavior happen mostly when the sum of ratings after an exclusion event is negative. In contrast, generalized approval or indifference (nor approval nor disapproval) of exclusion events favors a continuation of social exclusion (dashed line). When bystanders do not forcefully reward the inclusion in a particular round, offenders tend to exclude in the following round (dash-dots line); instead, when bystanders reward inclusion, it is perpetuated in the immediately following round (dotted line).

Figure 2.6: Potential offender's behavior depending on previous ratings



Notes: This figure includes the density for the social exclusion decisions of offenders in t and t+1, in terms of the sum of ratings from both bystanders in t.

Next, we estimate a logit regression for the social exclusion at t+1 depending on the feedback from bystanders at t. Table 2.6 presents the results for the estimations. Columns (1) and (2) regress on the sum of opinions of bystanders; columns (3) and (4) use the number of bystanders who punish the offender (0, 1, or 2); columns (5) and (6) include categories for the opinion of bystanders to potential offenders' actions.[14] We include controls for sociodemographic characteristics, personality traits, if subjects are finance or economics students, and a measure for inattention on the even columns. Here we see that if there is exclusion at t and bystanders reward the offender, the probability of exclusion in the following round increases. Once we analyze the results for the number of times bystanders punish a potential offender in period t, we find that if there is no exclusion at t, and if one of the bystanders punishes the offender

[14]We define that bystanders approve of an action when the sum of opinions is positive, while disapproval happens when the sum of opinions is negative. If the sum of opinions is zero, we say that bystanders are indifferent to potential offenders' actions

compared to none, this increases the probability of exclusion in t+1. In the presence of social exclusion at t, and with a single punishment decision, the probability of exclusion reduces in the following round. The results once we include the opinion of bystanders to potential offenders' actions indicate that if bystanders punish ostracism, social exclusion reduces in the following round, but that when bystanders are either indifferent or reward social exclusion, or when they punish inclusion, the exclusion increases the following round.

Table 2.6: Logit results for exclusion behavior after feedback

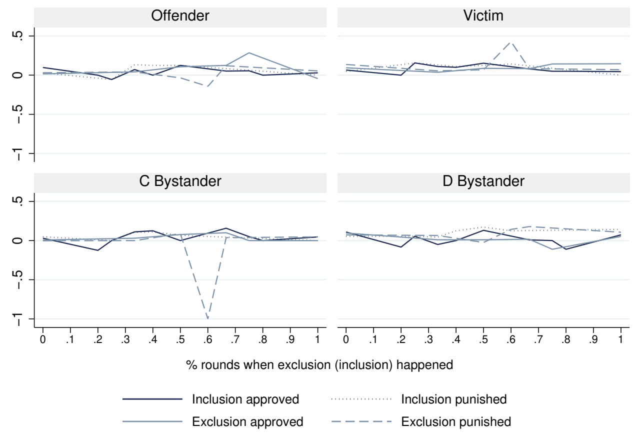
	Social exclusion in t+1					
	(1)	(2)	(3)	(4)	(5)	(6)
Social exclusion in t	-0.289	-0.564	2.143***	2.107***		
	(0.419)	(0.405)	(0.516)	(0.509)		
Sum of opinions of bystanders in t	-0.194***	-0.229***				
	(0.053)	(0.053)				
Social exclusion in t × Sum of opinions of bystanders	0.376***	0.402***				
	(0.107)	(0.110)				
One bystander punishes offender in t			2.114***	2.174***		
			(0.497)	(0.496)		
Two bystanders punish offender in t			-1.727*	-1.701*		
			(1.026)	(1.023)		
Social exclusion in t × One bystander punishes offender			-4.072***	-4.159***		
			(0.805)	(0.904)		
Indifferent to ostracism					1.128**	1.208**
					(0.527)	(0.601)
Social exclusion is rewarded					2.508***	2.456***
					(0.825)	(0.896)
Inclusion is punished					2.833**	2.947***
					(1.166)	(1.071)
Indifferent to inclusion					0.799	1.051*
					(0.563)	(0.621)
Inclusion is approved					0.144	0.184
					(0.439)	(0.479)
Constant	0.053	1.033	-1.227***	-1.155	-1.329***	-0.537
	(0.343)	(1.833)	(0.189)	(1.608)	(0.412)	(1.740)
Observations	276	276	276	276	276	276
Clusters	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes

Notes: Logit regressions. Dependent variable: social exclusion in t+1. Reference category is no exclusion of B players in columns (1) and (2), no exclusion and no punishment from bystanders in columns (3) and (4), and social exclusion is punished in columns (5) and (6). Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Result 6: Ratings from bystanders affect potential offenders' decisions. In particular, generalized disapproval of exclusion favors changes towards inclusion in the game, while generalized approval of inclusion maintains that behavior.

To understand the effects of punishment on our performance and emotional outcomes, we need to find a measure for the intensity of the exclusion during the second group stage from Part 2. We create a variable that combines the ratings from bystanders after the exclusion decisions of potential offenders. We first identify if a social exclusion event is generally approved or disapproved by both bystanders using the sum of their ratings in a given round. If the sum of ratings is positive, we say that bystanders approve of potential offenders' actions; if it is negative, we say that bystanders disapprove potential offenders' decisions. If the sum of ratings is zero, we say that, on average, bystanders are indifferent to potential offenders' social exclusion behavior. This classification gives us six possible categories: social exclusion is punished, social exclusion is rewarded, indifference to social exclusion, inclusion (no social exclusion) is punished, inclusion is rewarded, and indifference to inclusion. Since our performance and emotions outcomes are available for the part and not the round, we compute the number of times social exclusion (inclusion) is approved (disapproved) relative to the number of times social exclusion (inclusion) occurs during that group stage. Figure 2.7 shows the average changes in the normalized RET performance of subjects after the second group stage of Part 2, compared to the baseline results. On the x-axis, we have the percentage of rounds when social exclusion (inclusion) happened. We can observe that when ostracism is approved in more than 70% of the rounds where it happens, the RET performance of offenders appears to increase. In contrast, when it is punished in around 60% of rounds, the performance of offenders decreases. Oppositely, victims who observe punishment in around 60% of the exclusion events outperform themselves regarding baseline measures. When bystanders punish either decision of potential offenders, D players appear to have an increase in their performance. The same is not true for C bystanders.

Figure 2.7: Average changes in normalized RET by the reactions to social exclusion decisions, by role.



Notes: These figures present the average changes in the normalized RET performance of participants (y-axis). On the x-axis we include the percentage of rounds where either exclusion or inclusion happened. We present the by role and by the combination of potential offender actions and bystanders' reactions. As there is only a possibility for exclusion, these are potential offenders or victims.

A regression analysis further explores these findings by considering the interaction of social exclusion and punishment. Table 2.7 presents the OLS results for the normalized changes in RET performance between Part 1 and the second group stage of Part 2 for the potential offenders and the potential victims (analogous results for the bystanders are available in Table B.5 in the Appendix). In columns (1), (2), (5), and (6), we present the results using the number of times potential offenders get punished during the group stage as the independent variable and the categories of social exclusion previously defined. The rest of the columns use categories for punishment that use the median as reference (no punishment, punished once or twice, and punished more than twice, with $p_{50} = 2$). The even columns include the usual controls and control for the changes in the reported emotions between the second group stage from Part 2 and the baseline. The reference categories are no exclusion and no punishment.

Table 2.7: OLS results for within-subject changes in RET performance by punishment, for potential offenders and victims

	$\frac{RET_3 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender				*Potential Victim*			
Number of punishment events by part	0.017	-0.032			0.040***	0.086**		
	(0.031)	(0.041)			(0.013)	(0.038)		
Less than 3 exclusions by part	-0.048	-0.013	-0.307***	-0.210***	0.089	0.093*	0.147***	0.121
	(0.083)	(0.076)	(0.101)	(0.074)	(0.053)	(0.049)	(0.051)	(0.087)
3 or more exclusions by part	0.143**	0.151**	-0.403***	-0.188	-0.016	-0.076	0.158*	0.170
	(0.061)	(0.073)	(0.141)	(0.132)	(0.091)	(0.117)	(0.084)	(0.111)
Punishment events × Less than 3 exclusions	0.017	0.037			-0.054**	-0.101**		
	(0.048)	(0.048)			(0.021)	(0.045)		
Punishment events × 3 or more exclusions	-0.052	0.003			-0.022	-0.048		
	(0.034)	(0.047)			(0.030)	(0.045)		
Punished once or twice			0.005	-0.027			0.065***	0.153***
			(0.062)	(0.066)			(0.022)	(0.046)
Punished more than twice			0.375***	0.211*			-0.106*	-0.072
			(0.123)	(0.117)			(0.059)	(0.094)
Punished once or twice × Less than 3 exclusions			0.330**	0.253**			-0.152**	-0.203*
			(0.122)	(0.107)			(0.060)	(0.102)
Punished once or twice × 3 or more exclusions			0.495***	0.302**			-0.187**	-0.323**
			(0.151)	(0.137)			(0.092)	(0.141)
Constant	0.055	0.759*	0.057	0.727*	0.041*	0.026	0.041*	0.053
	(0.037)	(0.375)	(0.038)	(0.369)	(0.020)	(0.171)	(0.021)	(0.201)
Observations	69	69	69	69	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the second group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: changes in emotions (between the second group stage of Part 2 and the baseline), sociodemographic variables (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

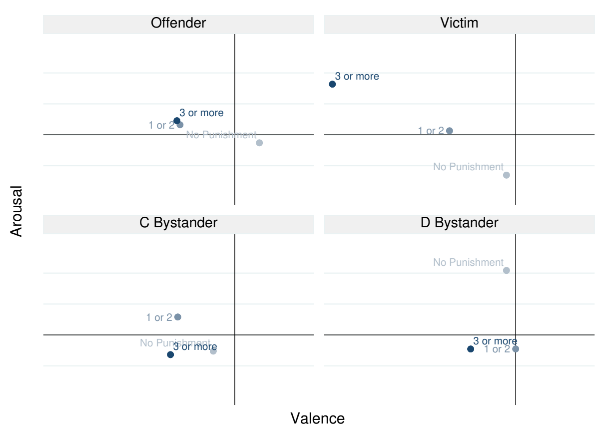
When the number of punishment events increases, the RET performance of potential victims increases with respect to the baseline. At the same time, it decreases for potential offenders, although this last result is not statistically significant. When we include the categories for punishment, we see that when offenders decide to exclude and are lightly punished (once or twice), they outperform themselves in terms of the RET ($p = 0.054$). The effect is larger for those that decide to exclude in three or more rounds ($p < 0.05$). The opposite happens to victims of exclusion that observe few punishment decisions of bystanders, but the effect is not significant. There are no important changes to be reported of bystanders.

Result 7: The number of times a potential offender is punished positively affects the RET performance of potential victims.

Result 8: When punishment to social exclusion is moderate (once or twice per group stage), the RET performance of offenders increases.

Figure 2.8 presents the changes in the two dimensions of the reported emotions from Part 1 to the second group stage in Part 2, using the punishment categories defined above. It appears that punishment negatively affects the emotions of all players. In particular, the valence of punished potential offenders and of potential victims who observe punishment reduces. The changes in the valence reported by potential victims is larger when punishment happens three or more times. C bystanders report lower valence of emotions compared to baseline measures, even in cases when there is no punishment. D bystanders, however, only report feeling relatively worse than during baseline when the punishment happens three or more times during the third group stage. In terms of arousal, the victims who observe above median punishment report intenser emotions, as do D bystanders in cases of no punishment.

Figure 2.8: Changes in the two dimensions of the reported emotions by punishment, by role



Notes: These figures present the changes between Part 1 (baseline) and the second group stage from Part 2 (punishment) for the two reported dimensions in the emotion test. We graph both dimensions by role for each of the defined categories for exclusion intensity (no exclusion, one or two, and three or more rounds). As there is only a possibility for exclusion, these are potential offenders or victims.

Now we move on to analyze the results in the reported emotions after the punishment possibility. Table 2.8 presents the results for within-subjects changes in the reported emotions for potential offenders and potential victims, interacting categories for punishment and exclusion. Columns (1) to (4) present the results for the changes in valence; columns (5) to (8) present the results for the changes in arousal. We include the usual controls in the even columns. Here we can observe that heavy punishment, compared to no punishment, decreases the valence dimension of emotions for potential offenders. The changes are larger and significant for those punished above the median. The results are not significant for the interaction of bystanders' punishment decisions with offenders' exclusion behavior. There are no general changes in the reported arousal levels of potential offenders after punishment. Potential victims reported valence is also negatively affected by above the median punishment, compared to no punishment to potential offenders. However, the reported arousal increases when potential offenders are lightly punished compared to the no-punishment situation.

Result 9: Above the median punishment negatively affects valence of potential offenders and potential victims compared to the no punishment situation.

Table 2.8: OLS results for within-subject changes in reported emotions by punishment, for potential offenders and victims

	$Valence_3 - Valence_1$				$Arousal_3 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Potential Offender								
Number of punishment events by part	-0.375	-0.392			0.000	-0.287		
	(0.304)	(0.328)			(0.625)	(0.847)		
Less than 3 exclusions by part	-0.532	-0.564	0.684	1.411*	0.286	-0.028	-0.947	-1.584**
	(0.662)	(0.806)	(0.784)	(0.791)	(0.631)	(0.647)	(0.843)	(0.649)
3 or more exclusions by part	-0.125	0.588	2.384**	3.213**	1.306*	1.538*	1.086	0.585
	(0.748)	(0.924)	(1.030)	(1.224)	(0.771)	(0.875)	(1.409)	(1.184)
Punishment events × Less than 3 exclusions	-0.003	0.128			-0.193	0.151		
	(0.409)	(0.414)			(0.671)	(0.846)		
Punishment events × 3 or more exclusions	0.308	0.211			-0.106	0.128		
	(0.336)	(0.380)			(0.646)	(0.887)		
Punished once or twice			-0.816*	-0.765			0.553	0.262
			(0.448)	(0.472)			(1.192)	(1.521)
Punished more than twice			-2.700***	-3.207***			0.300	0.807
			(0.942)	(1.010)			(1.131)	(0.944)
Punished once or twice × Less than 3 exclusions			-1.093	-1.687*			0.675	1.397
			(1.005)	(0.845)			(1.391)	(1.562)
Punished once or twice × 3 or more exclusions			-2.084*	-2.641*			-1.386	-0.567
			(1.223)	(1.351)			(1.866)	(1.882)
Constant	0.292	0.706	0.316	1.521	-0.000	1.170	-0.053	0.436
	(0.221)	(3.087)	(0.224)	(2.974)	(0.379)	(3.566)	(0.383)	(3.303)
Observations	69	69	69	69	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Potential Victim								
Number of punishment events by part	0.500**	0.006			1.406	1.542		
	(0.191)	(0.542)			(1.090)	(1.028)		
Less than 3 exclusions by part	-0.704	-0.618	-1.658	-1.598**	1.672**	1.550*	2.605**	3.167**
	(0.665)	(0.567)	(1.167)	(0.768)	(0.697)	(0.866)	(1.232)	(1.321)
3 or more exclusions by part	-2.168**	-2.091	-0.658	-0.785	0.093	-0.115	3.505**	3.793**
	(1.006)	(1.295)	(1.466)	(1.119)	(1.081)	(1.356)	(1.505)	(1.496)
Punishment events × Less than 3 exclusions	-1.017***	-0.509			-1.582	-1.696		
	(0.330)	(0.583)			(1.109)	(1.051)		
Punishment events × 3 or more exclusions	-0.537**	-0.095			-1.014	-1.125		
	(0.246)	(0.568)			(1.126)	(1.067)		
Punished once or twice			0.842***	-0.058			3.105*	3.095**
			(0.306)	(0.851)			(1.586)	(1.383)
Punished more than twice			-1.700	-1.584*			-1.200	-1.651
			(1.465)	(0.913)			(1.336)	(1.329)
Punished once or twice × Less than 3 exclusions			-0.251	0.543			-4.423**	-5.262***
			(1.233)	(1.004)			(1.982)	(1.812)
Punished once or twice × 3 or more exclusions			-2.342	-1.687			-6.505***	-6.992***
			(1.529)	(1.286)			(2.372)	(2.206)
Constant	0.167	1.557	0.158	2.492	-1.010*	-2.995	-1.105**	-3.207
	(0.300)	(2.186)	(0.306)	(2.204)	(0.497)	(2.639)	(0.499)	(2.741)
Observations	70	70	70	70	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variables are the changes in both dimensions of the reported emotions after the second group stage from Part 2 compared to the baseline. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9 presents the results for the same specifications from Table 2.8 but now considering both bystanders separately. The results for the reported emotions of the two bystander types go in the opposite direction. In particular, it seems that punishment negatively affects the valence dimension of emotions of C bystanders. Oppositely, punishment positively affects the reported valence of D bystanders.

Table 2.9: OLS results for within-subject changes in reported emotions by punishment, for bystanders

	$Valence_3 - Valence_1$				$Arousal_3 - Arousal_1$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C Bystander								
Number of punishment events by part	-0.938***	-0.607			-0.719	-0.380		
	(0.319)	(0.458)			(1.443)	(1.355)		
Less than 3 exclusions by part	-0.020	-0.202	0.421	-0.128	0.308	0.533	-0.237	-0.486
	(0.538)	(0.646)	(0.428)	(0.812)	(0.700)	(0.768)	(0.587)	(0.757)
3 or more exclusions by part	-0.798	-1.282	0.543	0.046	1.241	1.267	0.985	0.698
	(0.759)	(1.108)	(1.206)	(1.586)	(1.421)	(1.646)	(0.962)	(1.290)
Punishment events × Less than 3 exclusions	0.847**	0.489			0.552	0.118		
	(0.358)	(0.499)			(1.462)	(1.382)		
Punishment events × 3 or more exclusions	1.055***	0.780			0.486	0.142		
	(0.383)	(0.548)			(1.472)	(1.394)		
Punished once or twice			-1.579***	-1.221*			0.263	0.627
			(0.428)	(0.622)			(3.049)	(2.746)
Punished more than twice			-0.900*	-0.585			-0.500	-0.264
			(0.493)	(0.829)			(0.685)	(0.752)
Punished once or twice × Less than 3 exclusions			0.988**	0.957			0.464	0.405
			(0.471)	(0.846)			(3.101)	(2.775)
Punished once or twice × 3 or more exclusions			0.457	0.092			-0.385	-0.401
			(1.306)	(1.645)			(3.366)	(3.113)
Constant	-0.437	0.567	-0.421	0.533	-0.135	1.808	-0.263	1.164
	(0.421)	(3.192)	(0.428)	(3.268)	(0.461)	(3.580)	(0.451)	(3.272)
Observations	69	69	69	69	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes
D Bystander								
Number of punishment events by part	1.094***	1.039***			0.594	0.776**		
	(0.263)	(0.339)			(0.512)	(0.374)		
Less than 3 exclusions by part	-0.422	-0.389	-2.658	-2.370	-1.873**	-1.904**	-2.316**	-2.550**
	(0.650)	(0.668)	(1.897)	(1.758)	(0.769)	(0.899)	(0.894)	(1.212)
3 or more exclusions by part	-0.006	0.041	-1.058	-1.242	-2.577**	-2.017	-2.316	-2.161
	(0.786)	(0.745)	(1.971)	(1.827)	(1.145)	(1.368)	(1.431)	(1.658)
Punishment events × Less than 3 exclusions	-1.315***	-1.341***			-0.523	-0.756		
	(0.310)	(0.376)			(0.607)	(0.509)		
Punishment events × 3 or more exclusions	-1.123***	-1.210***			-0.358	-0.639		
	(0.310)	(0.399)			(0.546)	(0.464)		
Punished once or twice			1.842***	1.757***			1.184	1.350**
			(0.274)	(0.402)			(0.738)	(0.511)
Punished more than twice			1.000	0.431			0.800	0.978
			(1.800)	(1.628)			(1.170)	(1.494)
Punished once or twice × Less than 3 exclusions			0.476	0.135			-0.593	-0.690
			(1.915)	(1.817)			(1.114)	(1.281)
Punished once or twice × 3 or more exclusions			-0.942	-0.464			-0.984	-1.252
			(2.034)	(1.896)			(1.724)	(1.697)
Constant	0.177	2.273	0.158	1.077	1.344***	0.808	1.316**	0.385
	(0.269)	(2.137)	(0.274)	(2.262)	(0.481)	(2.986)	(0.485)	(3.047)
Observations	70	70	70	70	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variables are the changes in both dimensions of the reported emotions after the second group stage from Part 2 compared to the baseline. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Result 10: The position C bystander occupies concerning potential offender and potential victim seems to be generating negative emotions, while the position D bystander occupies leads them to report positive emotions.

2.4.3 Role Reversal

Now we turn to study the behavior of former potential victims and potential offenders once there is role reversal. We begin our analysis by exploring the behavior of potential offenders depending on how they were treated as victims in terms of exclusion. Table 2.10 presents the results for the OLS estimation of the number of times a former victim, now in the role of potential offender, decides to exclude participants in role B, depending on the number of times she was excluded in Part 2. Column (1) presents the results for the estimation without including controls. From Column (2) onward, we include sociodemographic variables, if the subject is an economics or finance major, and a measure of intention as controls. In Column (3), we add the changes in the reported emotions when they had the role of victims after the possibility of exclusion, compared to the baseline measures. Here we want to explore if the change of emotions we created when subjects were in the victim position explains this result. In Column (4) we include the personality index.

Table 2.10: Exclusion behavior of former potential victims in the role of potential offender

	Number of exclusion events in Part 3			
	(1)	(2)	(3)	(4)
Number of times excluded as potential victim	0.359***	0.413***	0.368***	0.382***
	(0.108)	(0.117)	(0.116)	(0.112)
$Valence_2 - Valence_1$			-0.078	-0.063
			(0.093)	(0.088)
$Arousal_2 - Arousal_1$			-0.110	-0.072
			(0.097)	(0.088)
Personality index				-0.987
				(0.726)
Constant	1.442***	0.971	0.958	0.817
	(0.359)	(1.472)	(1.605)	(1.414)
Observations	70	70	70	70
Clusters	35	35	35	35
Controls	No	Yes	Yes	Yes

Notes: OLS regressions. Dependent variable is number of times a potential offender excludes in Part 3. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Although it is clear in the instructions that potential victims are not the necessarily the same players that excluded the now potential offenders, we observe that the social exclusion decision of former potential victims is positively and significantly associated with the number of times they were excluded from the game in Part 2. Positive changes in valence and arousal when they were in the potential victim role and after observing the possibility for social exclusion decrease the number of times a potential offender decides to exclude, but these results are not statistically significant. Neither is our measure for personality traits.

Result 11: Former victimization increases retaliation against someone that is not necessarily at fault for former exclusion. This result does not appear to be driven by the emotions generated after being a victim of exclusion or personality traits.

Last, Table 2.11 presents the results for participants in the roles of potential offenders and victims in Part 3. These participants had the role of potential victims and potential offenders during parts 1 and 2, respectively. Our dependent variable is the within change in the RET performance after the first group stage from Part 3 and the baseline, normalized with the baseline performance to account for individual ability. We include as independent variables the intensity of exclusion in both parts in levels and using categories (with no exclusion as the omitted category). Odd columns include controls. We see that participants in the role of offender that decide to exclude see a reduction in their RET performance, compared to those that decide to not exclude at all. The changes in RET are higher for those that heavily exclude B players than for those that exclude once or twice. The results improve for those that decide to exclude three or more times when they are offenders, and that were excluded as victims.

Result 12: Former victims that decide to exclude once they are in the offender position observe a negative change in their performance compared to those that do not exclude. However, the performance increases for those that decide to intensely exclude that were also excluded as victims.

Table 2.11: OLS results for within-subject changes in RET, role reversal

	$\frac{RET_4 - RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Potential Offender				*Potential Victim*			
Number of exclusion events in Part 3	0.008	-0.004			0.036	0.044		
	(0.018)	(0.019)			(0.024)	(0.028)		
Number of exclusion events in Part 2	0.004	0.018			0.033	0.038		
	(0.016)	(0.018)			(0.026)	(0.031)		
Exclusion events Part 3 × Exclusion events Part 2	-0.006	-0.004			-0.011	-0.014		
	(0.006)	(0.006)			(0.008)	(0.010)		
Less than 3 exclusions in Part 3			-0.011	-0.026			-0.174	-0.184
			(0.102)	(0.099)			(0.119)	(0.148)
3 or more exclusions in Part 3			-0.154***	-0.217**			0.271***	0.314**
			(0.041)	(0.082)			(0.074)	(0.137)
Less than 3 exclusions in Part 2			-0.105	-0.119			-0.101	-0.067
			(0.078)	(0.072)			(0.081)	(0.104)
3 or more exclusions in Part 2			-0.154***	-0.121**			0.037	0.104
			(0.041)	(0.059)			(0.074)	(0.135)
Less than 3 exclusions in Part 3 × Less than 3 exclusions in Part 2			0.034	0.081			0.313**	0.283
			(0.132)	(0.129)			(0.147)	(0.176)
Less than 3 exclusions in Part 3 × 3 or more exclusions in Part 2			0.108	0.168			0.225*	0.171
			(0.107)	(0.105)			(0.133)	(0.187)
3 or more exclusions in Part 3 × Less than 3 exclusions in Part 2			0.221**	0.290***			-0.162*	-0.229
			(0.088)	(0.104)			(0.095)	(0.164)
3 or more exclusions in Part 3 × 3 or more exclusions in Part 2			0.207***	0.277***			-0.264**	-0.363***
			(0.058)	(0.080)			(0.105)	(0.116)
Constant	0.102***	0.033	0.154***	0.100	-0.010	0.546	0.063	0.567
	(0.037)	(0.214)	(0.041)	(0.223)	(0.067)	(0.493)	(0.074)	(0.544)
Observations	70	70	70	70	69	69	69	69
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the first group stage from Part 3 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion in Part 2 and no exclusion in Part 3. Controls: sociodemographic (age, female, number of siblings, number of household members, socioeconomic status), changes in emotions after possibility of exclusion in Part 2 with respect to baseline, personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Participants in the victim role who first decided as potential offenders outperform themselves concerning baseline measures whenever they are heavily excluded (more than twice), regardless of how much they excluded when they were in the offender position.

Result 13: Former offenders that are heavily excluded do not see their performance affected and outperform themselves.

2.5 Discussion and conclusion

We propose a laboratory experiment to study social exclusion. This particular manifestation of peer aggression is difficult to identify and discourage. Thus, understanding its dynamics can help inform policies to reduce the prevalence of this phenomenon and the negative consequences it generates. We are interested in studying the performance and emotional effects of exclusion. In addition, we want to see the capacity of bystanders to reduce the incidence of this phenomenon and the negative consequences it has on the direct victims of the exclusion.

We find that social exclusion events do not generally affect the performance of the victims or the bystanders as we expected. The older non-excluded victims and the heavily excluded participants outperform themselves compared to baseline, but this last result is not statistically significant. Using an incentivized task to measure performance may be causing subjects excluded in three or more opportunities to increase their effort during the task to try to compensate for the losses from social exclusion. Maybe only those for whom the psychological costs are large enough are willing to incur the costly effort required to increase productivity. Lightly excluded participants have a smaller performance change than non-excluded B players. The emotions capture the effect that the exclusion generates in them. The results in terms of the reported emotions are more consistent with what we expected. Victims of the exclusion and participants in the position of an offender that decides to include B players in all rounds report feeling significantly worse than baseline. This last result for victims is consistent with the literature that uses an ostracism manipulation, even though the incentives to exclude participants are common knowledge to all. Oppositely, offenders in three or more rounds report feeling relatively better than those who did not exclude.

Although we do not find results on emotions and performance for bystanders as they do in the psychology literature, we observe that bystanders negatively rate exclusion from offenders. This behavior can be seen as empathy-related, even when punishment is not costly for them. In addition, participants in the offender position change their conduct in response to the signals sent by the bystanders. As a result, social punishment effectively reduces the incidence of exclusion. Another highlight of punishment is that victims outperform themselves compared to baseline. These findings hint at the importance of bystanders actively intervening against the exclusion. However, even when there are no monetary costs of punishment for most participants, forcing peers to intervene appears to be causing an emotional toll on potential offenders and victims and the bystanders closer to the victims. On the other hand, the bystanders farther from the victims report feeling relatively better than baseline. Although both bystander types are statistically balanced in terms of observables, they are impacted differently by the game dynamics. This difference in results may have to do with the closer distance between victims and C bystanders and with the fact that they are the ones receiving the ball whenever offenders decide on exclusion.

Last, we explore the effects of being a former victim regarding future victimization behavior and changes in performance. We find that the possibility of excluding when they were previously ostracized increases the RET performance of former victims. This result contrasts with the lack of effect from victimization on the RET performance of the original offenders. We also find that the higher the number of times a person was excluded as a victim, the higher the number of times she decides to exclude other participants. We observe negative reciprocity even when experiment instructions mention that they were not interacting with the person who was their potential offender in a previous game stage. This dynamic may suggest how these behaviors become socially prevalent and socially accepted exclusion rules arise.

Chapter 3

Leaders and communal norms in the workplace

3.1 Introduction

There is a gender imbalance in the time use of men and women in research, with women spending less time on research and more on teaching and service (see Taylor et al. (2006); Harter et al. (2011); Manchester and Barbezat (2013); Buckles (2019) for economics). In addition, more than men, women volunteer and are asked to volunteer for non-promotable tasks (Babcock et al., 2017). This is in spite of the reward structure, particularly for tenure-track faculty, which is often tied to performance in research. This metric does not consider tasks that are also key to the organizational objectives and are more communal in nature.

Part of the time-use differences may be explained by women having preferences closest to a communal norm than men. There is evidence that women are more altruistic (Brañas-Garza et al., 2018) and generally more generous than men (Doñate-Buendía et al., 2022). Similarly, women tend to be less efficiency-oriented and more likely to aim for equality than men (see Niederle (2020) for a review) and to choose egalitarian payoffs (Kamas and Preston, 2015). In contrast, men's other-regarding decisions appear to be context-dependent (Niederle, 2020), and they are more likely to choose competitive payoffs (Kamas and Preston, 2015). One possible alternative to accommodate communal tasks in the workplace and increase their valuation is encouraging an institutional shift towards a more communal norm.

In this paper, I study how leaders advocating for a communal norm encourage its approval by team members, depending on the gender of the leader. In addition, I study whether a meritocratic selection of the leader improves the acceptance of the communal norm. Last, I explore whether the gender composition of the group affects the endorsement of the communal norm suggested by the female or male leader. To do so, I propose an experimental design in which subjects interact in groups of three members. One group member is selected as the leader and decides on the form of payment for the relative performance of the group in a real-effort task (encrypting letters into numbers). One payment alternative implies ex-ante egalitarian earnings for the members of both groups (communal norm). The other payment alternative allows for potential gains from a tournament-style payment (agentic norm) in which the top-performing group earns a premium while the other group earns zero.

I use the strategy method to explore the approval of both norms by the group members who are not assigned to the role of the leader. The two non-leaders in each group have veto power over the leader's decision only when both agree against it. I randomly define a female or male leader and the gender composition of the group. In addition, I include two treatments to identify whether there are differences in group members' behavior based on why the leader is chosen: meritocratic or non- meritocratic. My design incorporates five additional decision stages that provide an individual measure of other-regarding preferences, individual preferences for a tournament environment without and with externalities, and group preferences for a communal vs. an agentic norm, with and without information on the gender composition of the groups.

I contribute to the three branches of the literature. First, I contribute to the branch that discusses that closing the gender gaps in labor outcomes requires "fixing the institutions" (Niederle, 2020). This literature has proposed several institutional changes that mainly respond to gender gaps in competitiveness and performance under competition.[1] However, the gender gaps in competitiveness may also be responding to

[1]These institutions include single-sex tournaments, use of stereotypical female tasks, feedback on relative performance, gender quotas or preferential treatment, team competition, and choice of the gender of the opponent (see Niederle (2020); Markowsky and Beblo (2022); Coffman and Roth (2021)). Recent literature has identified that some successful mechanisms to reduce the gender gaps in competitiveness are principals selecting competition on behalf of an agent (Fornwagner et al., 2020), priming subjects with power before competition (Balafoutas et al., 2018), assigning roles before competing (Zhang et al., 2020), or informing participants about the gender gap in competitiveness

differences in prosocial preferences of women and men.[2] As a result, I focus on the institutional changes that may increase the valuation of communal (other-regarding) traits in the workplace. This is in line with the responses of 10,000 women to the question "What needs to change so more women can reach their full potential?". In their report, BSR (2021) find that 80% of the surveyed women mention how discrimination and harmful gender stereotypes bound women to specific jobs and create an undervaluing of the characteristics typically associated with women.[3] Next, I contribute to the literature on leadership and gender. In a review, Shen and Joseph (2021) discuss that female leaders have a more democratic (instead of autocratic) leadership style than men and are more often classified as transformational leaders. Female directors are also more "benevolent and universally concerned" than their male colleagues (Adams and Funk, 2012). In addition, firms with more women in leadership roles have a smaller wage gap and offer equality in wages to new employees (Tate and Yang, 2015). Last, female policy makers in India increase the provision of public infrastructure directly related to women's preferences Chattopadhyay and Duflo (2004). Using a collective action game, Grossman et al. (2015) find that female leaders tend to be more cooperative than men, particularly in single-gender treatments. In contrast, female followers are more cooperative if there is no information about the leader's gender. Last, I contribute to the literature on gender composition of teams and the social impacts of gender diversity in the boardroom. There is evidence that female-majority groups give more to a recipient in a dictator game and choose an egalitarian division more frequently than male-majority groups (Dufwenberg and Muren, 2006). In addition, female board representation positively impacts social performance (corporate social responsibility),

and its implications on earnings (Kessel et al., 2021).

[2]Prosocial preferences appear to affect selection into competition (Balafoutas et al., 2012; Bartling et al., 2009; Dasgupta et al., 2019) and tournament performance (Balafoutas et al., 2012). Recent literature has found that introducing social incentives (percentage of earnings donated to a charity) increases women's performance in tournament payment schemes, but not men (Drouvelis and Rigdon, 2022). In addition, the option to share the rewards with the losers increases women's willingness to compete (Cassar and Rigdon, 2021b) and tournament performance (Cassar and Rigdon, 2021a). In the field, Houser and Schunk (2009) find that competition negatively affects the altruistic choices of boys, but not of girls, while Cassar and Zhang (2022) find that introducing voucher payments for things that matter to women (child-benefitting or gender-stereotypical goods) increases competitiveness in some cultures.

[3]Social role theory explains the existence of gender stereotypes and expectations arising from gendered labor divisions, not biological traits. As women bear and rear children, they develop nurturing traits that associate them with stereotypes of communality. Men's participation in the labor market has required them to develop skills such as competitiveness and self-promotion, associated with stereotypes of agentic behavior (Rudman and Glick (2021), Chapter 1).

ethical behavior, gender diversity at other organizational levels (see Kirsch (2018) for a review and Byron and Post (2016) for a meta-analysis),[4] and environmental performance of firms (Lu and Herremans, 2019).

In light of the evidence in the literature, I hypothesize that female leaders prefer the egalitarian condition more than male leaders. In addition, I anticipate that female majority groups will favor changes towards a more egalitarian rule. I also expect that female majority groups will shape the group's decision-making so that the preferences of female leaders are more frequently accepted by all. An empiric question remains on the potential differences in results when leaders are the people in the group with relatively higher performance.

I find that women, more than men, prefer an egalitarian distribution. This is true when they decide as individuals (stages 1 to 3), as part of a group (stages 4 and 5), and as group leaders (stage 6). The individual decisions appear to be driven by women being more risk-averse than men. As part of a group and as leaders, women tend to select the more communal distribution norm, even after controlling for their beliefs, performance, tournament preferences, and altruism. The gender composition of the group does not consistently favor one norm over the other. Women in female-majority groups are less likely to select an agentic norm and to change from an individual preference for equality to an agentic group norm. This is only the case for the most altruistic men in either group configuration. In addition, non-leaders seem more responsive to male leaders, so women appear less influential than men in favoring changes toward an egalitarian rule when they are in a leadership position.

The rest of the paper is organized as follows. In Section 2, I describe the experimental design and procedures. Then, in Section 3 I present the results. In Section 4, I discuss results and breakdown conclusions.

[4]Using an online business simulation game, Apesteguia et al. (2012) find that three women teams invest significantly more than other gender configurations in social initiatives.

3.2 Experimental design and procedures

3.2.1 Experimental design

I propose an experimental design consisting of an individual task followed by six decision stages (see Figure 3.1). I follow Chakraborty and Serra (2022)'s method for revealing subjects' gender. At the beginning of the experiment, I ask participants to complete a short questionnaire including their self-reported gender,[5] age, their major, whether they are pursuing an undergraduate or graduate degree, and their participation in previous experiments. I use the self-reported gender to condition a list of predetermined names they can choose from at a later stage.[6] Then, subjects participate in a Real Effort Task (RET) consisting of encoding three-letter "words" into numbers (Benndorf et al., 2019). In this task, participants must identify which three-digit number corresponds to each of the three letters forming the "word" according to an equivalence table with all the letters of the alphabet randomly organized (see Figure C.1 in the Appendix). The letter-to-number equivalencies, the order in which they appear, and the three letters to encode change randomly as each "word" is successfully completed. The encoding task lasts for two minutes, after which they are reminded of the total number of correct encryptions, and is preceded by a trial stage of the same time length. The performance in this task conditions the potential earnings from the experiment. After stages 3, 5, and 6, I include a series of belief elicitation procedures to be detailed below.

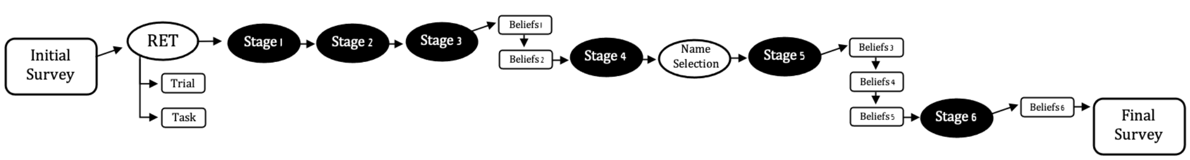


Figure 3.1: Timeline of the experiment

Once the RET concludes, participants make decisions in six distinct stages regarding their preferred form of payment for their task performance between two alternatives (see Table 3.1 for a summary). One of these stages is randomly selected at the end of

[5]Using the participant lists, I identified that two female participants self-selected as men. I include them as such in the regressions.

[6]I select the most common names for boys and girls (26 in each case) over the last 15 years as reported by official Colombian government sources.

the experiment to determine the experimental payments. In all stages, participants are part of a group of three people attending the session. The group configuration from stages 1 to 4 is entirely anonymous to everyone. In stages 5 and 6, I reveal information about the sex of the participants with whom they interact by giving the names they selected from the predefined list of names for each gender. These names uniquely identify each subject in the session and depend on the auto-reported gender in the initial survey. The computer randomly defines group configurations at the beginning of stages 1 and 4. Groups also change before Stage 5 with a procedure that maximizes the probability of mixed-gender (instead of single-sex) composition and of no tie in the first position (useful in Stage 6 for one of the treatments). From Stage 4 onward, the newly formed groups are paired with another group of three people also present in the session (supergroups) as a way to instrument for group interaction.

The first decision enables for the classification of subjects according to their prosocial preferences (altruism). Choosing the first alternative in Stage 1 allows all group members to earn 1.5 Experimental Currency Units (ECUs) per correct encryption during the RET. Oppositely, choosing the second alternative guarantees 4.5 ECUs per correct encryption for the Decision-Maker (DM), while the other two group members earn zero. Consequently, a subject selecting Alternative 1 is classified as an ex-ante egalitarian. In contrast, if she chooses Alternative 2, she is classified as self-serving. If this stage is selected to determine the experimental earnings, the decision of one randomly determined group member decides the payments for all three.

Stage 2 changes the decision environment by abstracting from other-regarding considerations (no externality condition) and introducing a tournament-style payment.[7]. Here, the first payoff alternative implies that the DM obtains 1.5 ECUs per correct encryption, while the second payoff alternative allows her to obtain 4.5 ECUS per correct answer, only if she is the top performer of her group, and zero otherwise. Since this decision does not affect the other group members, I gather information on individual preferences for a sure versus an uncertain payment that involves the thrill of competition. In Stage 3, I include the possibility that the selection of a tournament

[7]Although my experimental design does not include the typical approach to measure competition in an experimental setup and does not address changes in performance under competition, it does allow for a tournament-like payment that leads to unequal allocations. This design decision responds to the fact that my research question is related to prosocial preferences and not to performance in competitive environments.

compensation scheme affects others than the DM (externality condition). Unlike Stage 2, during Stage 3, the computer selects the decision of one group member to determine the revenues for all. In this case, the first payoff alternative remains the same as in Stage 1. However, the second payoff alternative implies that, if selected, the top performer of the group, and not only the DM as in Stage 2, will earn 4.5 ECUs per correct encryption, while the other group members will earn zero. Consequently, someone that chooses Alternative 2 in Stage 2 but chooses Alternative 1 in Stage 3 enjoys the thrill of the tournament but does not want to impose a negative externality on others.

Table 3.1: Alternatives per stage

		Alternative 1	Alternative 2
Individual Decision	Stage 1	1.5 ECUs per correct encryption for all group members	4.5 ECUs for DM, 0 for the other two group members
	Stage 2	1.5 ECUs per correct encryption for DM	4.5 ECUs for DM if DM is top performer, 0 otherwiwse
	Stage 3	1.5 ECUs per correct encryption for all group members	4.5 ECUs for group's top performer, 0 for the other two group members
Group Decision	Stages 4 and 5 (majority rule)	1.5 ECUs per correct encryption for all supergroup members	4.5 ECUs for top performing group, 0 for the other group
	Stage 6 (leader's choice)	1.5 ECUs per correct encryption for DM	4.5 ECUs for DM if DM is top performer, 0 otherwiwse
	Stage 6 (non-leader's choice)	Accept/Reject leader's choice of Alternative 1	Accept/Reject leader's choice of Alternative 2

Notes: ECUs stand for Experimental Currency Units (1 ECU corresponds to COP 1000; DM stands for Decision Maker. The top performer of a group is the group member with the highest number of correct encryptions during the task. The top performing group is the group in the supergroup with the highest sum of encryptions from all its members. In the case of a tie, the computer chooses the winner at random. Stages 4 and 5 have the same payment structure; the only difference is in the information of the group composition by gender.

In Stage 4, the participants are randomly assigned to a new group of three people who have to interact with another group of equal size. Here the decision is the same as in Stage 3, but now Alternative 2 deems the tournament winner to be the group with the highest sum of encryptions of its members.[8] This mimics a situation in which

[8]This payment scheme, different from the revenue-sharing scheme in which profits are divided equally, avoids the emergence of an additional consideration related to the desire or not to share

team members specialize in different tasks, but the team outcome depends on the sum of individual efforts. Although everyone states their preference on how to pay the members of both groups, the group's decision depends on what the majority chooses. Analogously to Stage 3, the computer randomly selects the decision of one of the groups belonging to each supergroup to determine the earnings.[9] In Stage 5, participants are randomly assigned to a new group/supergroup that remains the same until the end of the experiment. Here, the decision is the same as in Stage 4, with the only difference being the information on the group's gender composition.[10] Last, the decision set in Stage 6 depends on the role to which a subject is assigned: leader and non-leader. Group leaders have to make the same decision as in Stage 5. The non-leaders must decide if they want to accept the leader's decision. I use the strategy method for non-leaders, so I recover their preferences for the leader's choice of alternatives 1 and 2. Here, the majority rule no longer applies. The leader's decision determines the group's preferences unless both non-leaders reject the alternative chosen by her.

Asides from the within-subject treatments presented above, my experiment has a between-subject design with two different types of sessions that will allow me to evaluate the effect of meritocracy by comparing merit vs. random leader assignment. In some sessions, the group leaders are randomly selected, while in others, the group leaders are the person within the group with the best relative score. The experiment ends with a questionnaire that collects sociodemographic information (including parents education), a general hypothetical risk question ("How do you see yourself: are you, in general, a person who is fully prepared to take risks, or do you try to avoid them?") that predicts individual risk attitudes (Dohmen et al., 2011) and others about risk in specific contexts, and the Big Five agreeableness questions that correlate negatively with leadership ambition and predict transformational leadership (see discussion in (Chakraborty and Serra, 2022)).

profits with other members of the group. Thus, I isolate the effect of perception concerning whether or not they believe their group can win, subject to their risk aversion.

[9]If the majority of the selected group decides on the first alternative, all six participants obtain 1.5 ECUs per correct encryption. If the selected group decides on the second alternative, the members of the top performing group, as given by the sum of correct encryptions of all of its members, earn 4.5 ECUs per correct encryption, and the members of the other group earn zero.

[10]By design, both groups of the supergroup have the same gender composition except for four supergroups (24 participants) due to the gender distribution of the participants who arrived at their particular sessions.

3.2.2 Belief elicitation tasks

In addition to the decisions throughout six stages, I include incentivized elicitation[11] of beliefs in three parts of the experiment: after stages 3, 5, and 6 (see Figure 3.1). At the end of Stage 3, I ask participants their believed position within their group, and, separately, I ask them to rank themselves among all the participants in the session. A correct elicitation of beliefs grants the participants 1 ECU in the first case. The second elicitation of beliefs is paid according to the distance from their actual position in both directions, with a maximum payment of 3 ECUs and a minimum of zero. The second elicitation of beliefs takes place after Stage 5. In this case, I ask participants to rank their group compared to the other group in the supergroup (with information on group gender composition) and a third anonymous randomly selected group in the session. For each correct ranking, they can earn 1 ECU. I then ask them the full rank of the three members of their group according to their performance in the task, with each correct guess adding 0.5 ECUs to their earnings. In addition, I can identify possible gender stereotypes associated with the task by observing how they rank group members. Finally, at the end of Stage 6, I elicit beliefs about the leader's decision (for the non-leaders) with an incentive of 1 ECU, and of the acceptance or rejection decisions by the non-leaders (for the leaders), which adds 0.5 ECUs for each correct guess. The elicited beliefs are used as controls in the regressions and for analyzing selection into a tournament-style compensation.

3.2.3 Experimental procedures

I carried out twelve experimental sessions at the Rosario Experimental and Behavioral Economics Lab (REBEL) from Universidad del Rosario, six per treatment, between March 15th and 18th, 2022. Each session consisted of 18 participants, with two sessions (one per treatment) with 12 subjects. The experiments were programmed in zTree (**?**), and I manually recruited subjects from REBEL's databases.[12] Once the experiment started, I sequentially handed participants physical copies of the General, RET, and Stage 1 instructions. I read aloud the General Instructions and a summary of the RET and Stage 1 instructions. The rest of the instructions were available on the

[11] I define small monetary incentives to reduce the hedging opportunities (Blanco et al., 2010).

[12] As a consequence of this manual procedure, one subject participated in two sessions. I eliminated her second participation from the final sample. I kept the rest of the subjects with whom she was paired since there was no actual interaction among them.

experimental screens. Participants spent an average of 42 minutes in the activity and received an average payment of COP 30.000 (about $9), including a show-up fee. At the end of the experiment, they received feedback on the Stage randomly selected by the computer for payments and on the results from the belief elicitation procedures. No feedback occurred elsewhere in the experiment. I collected banking information and transferred earnings within 24 hours.

3.3 Results

I study the decisions of 203 subjects. Table 3.2 presents the summary statistics by gender. I include the main sociodemographic characteristics (age, the number of siblings by gender, a proxy for income (strata), and a dummy that takes the value of one when the education of the mother is above that of the father), a dummy that takes the value of one for economics or finance majors, the trial and task RET performance, and other information, including the risk measure, an agreeableness index[13], a proxy for other-regarding preferences using the percentage of participants that choose Alternative 1 in Stage 1, and the selection in the merit treatment. These variables are used as controls for understanding individual (stages 1 to 3) and group decisions in anonymous and gendered settings (stages 4 and 5, respectively) and as leaders or non-leaders (Stage 6). Overall, there is balance in these variables across initial roles, as shown by the p-values from the mean comparison tests. The agreeableness index is the only variable for which genders do not appear balanced.

[13] I use the answers to the questions in the agreeableness dimension from the Big Five and build an index using a Confirmatory Factor Analysis (CFA).

Table 3.2: Summary statistics and mean comparison tests, by treatment

	Mean		p-value
	Men (N=100)	Women (N=103)	Men vs Women
Sociodemographic Characteristics:			
Age	21.10	21.24	0.312
Number of female siblings	0.81	0.90	0.117
Number of male siblings	0.79	0.70	0.112
Strata	3.61	3.53	0.149
Mother's level of education above of father's	0.39	0.34	0.068
University information:			
Economics or finance	0.50	0.30	0.068
RET results:			
Trial result	6.96	6.85	0.243
Trial mistakes	0.21	0.20	0.088
Task result	8.43	8.19	0.217
Task mistakes	0.19	0.19	0.069
Other			
Participation in previous experiments	0.90	0.82	0.049
Preference for equal payment conditions	0.30	0.49	0.068
Willingness to take risks	3.31	2.51	0.159
Agreeableness index	-0.03	0.03	0.023
Merit treatment	0.44	0.56	0.070

Notes: This table shows the mean by gender for the final survey variables and the results in the RET. It also includes the p-value for the mean differences between gender. I present results for age, number of siblings by sex, and socioeconomic status following the Colombian classification. I also include a dummy that takes the value of one when the mother's schooling level is above that of the father's and another dummy for Economics or Finance majors. In addition, I present the performance during the encryption task and the frequency of the participation in previous experiments. I also include the percentage of participants choosing Alternative 1 in Stage 1 as a proxy for other-regarding preferences. I build an index factoring the nine questions for the agreeableness dimension of the Big Five personality test and present the results for the question on risk preference. Last, I show the percentage of participants in the merit treatment.

3.3.1 Individual decisions

According to their decisions in Stage 1, 39.4% of participants are classified as ex-ante egalitarian. This is the case for 30% of men and 48.45% of women ($\chi^2 = 7.31$ and $p < 0.01$). This result is consistent with the literature on gender differences in egalitarian preferences (see Niederle (2020) for a review). In addition, 43% of men and 28.16% of women select the tournament alternative when this decision does not impose a negative externality on others (Stage 2)[14] ($\chi^2 = 4.89$ and $p = 0.027$) Once I include the possibility of causing a negative externality in the tournament compensation (Stage 3), both men and women move towards the egalitarian alternative (84.5% of women and 68% of men; $\chi^2 = 7.62$ and $p < 0.01$). I classify participants according to their decisions in stages 1 and 2 I find that 37% of women in my sample are egalitarian that prefer not to subject themselves to a tournament. At the same time, 35% of women are not egalitarian that prefer not to select a tournament compensation. Oppositely, 35% of men are not egalitarian and dislike tournaments, while 35% are not egalitarian and prefer to select a tournament compensation. In terms of subjects' decisions in stages 2 and 3, I find that 65.1% of women never choose the tournament option. In contrast, 23% of men always select the tournament, and 20% of them opt out from the tournament when it creates a negative externality on others.

Result 1: Women are more likely to select the ex-ante egalitarian alternative.

Result 2: Women are less likely to choose the tournament compensation scheme, with or without externality, than men.

To analyze the reasons for gender differences for selecting a tournament, I begin exploring the performance of women and men in the RET. Men correctly encrypt 8.43 words while women encrypt 8.19 on average. This difference is not statistically significant using a two-sided t-test ($p = 0.279$). In addition, both women and men make 0.19 mistakes ($p = 0.952$). Figure 3.2 shows the Cumulative Distribution Function of the number of correct encryptions for women and men. Given there is no gender gap in performance, I should not expect to observe gender differences in the probability of winning a tournament.

[14] In a similar decision environment but using a different task, Niederle and Vesterlund (2007) find that 55% of the men and 25% of the woman submit their piece-rate performance to a tournament.

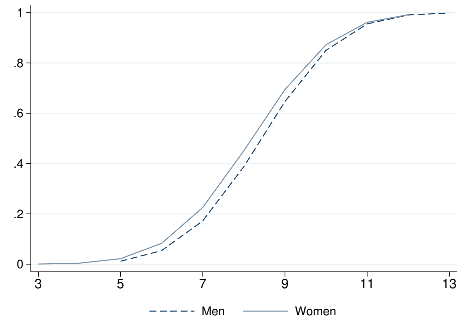


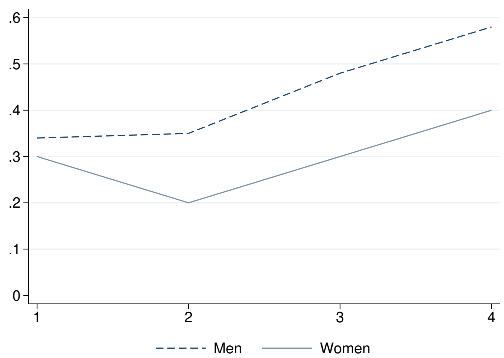
Figure 3.2: Normal CDF of performance in encryption task, by sex

Furthermore, women who decide to participate in the tournament are not significantly different in terms of RET performance from those that decide not to. In the absence of externality, women participating in the tournament correctly encrypt 8.34 words while those deciding against the tournament encrypt 8.12 words (a bilateral t-test generates a $p = 0.461$). Comparing the results in the presence of an externality, the number of correct encryptions are 7.93 and 8.24, respectively ($p = 0.484$). Conversely, men who decide to participate in the tournament without externality solve, on average, 8.72 words, while those who opt-out from the tournament correctly encrypt 8.21 words. In the presence of negative externalities from opting into a tournament, men willing to compete encrypt 9.06 words while those who do not encrypt 8.13 words. In both cases, men who opt into the tournaments have a higher performance in the task than men who do not ($p = 0.095$ and $p < 0.01$, respectively). These results are consistent with Niederle and Vesterlund (2007) for the decision to submit the piece rate performance to a tournament (their Stage 4).

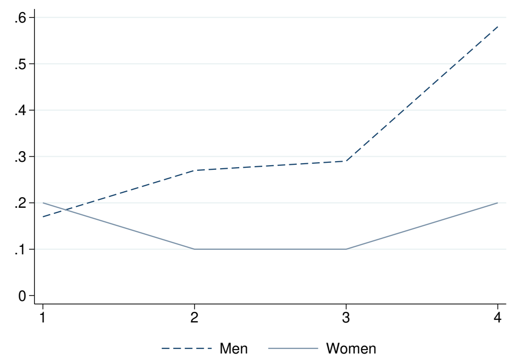
Result 3: Men and women are no different in terms of performance in the task. However, men choosing a tournament compensation have a relatively better performance than men choosing the egalitarian alternative. This is not the case for women.

Figure 3.3 shows the proportion of women and men who select the tournament without or with externality (Figures 3.3a and 3.3b, respectively) by performance quartile, with one being the worst and four the best quartile. For every performance level, the selection of a tournament compensation is more common in men than women. Among participants in the highest quartile (10 correct encryptions or more), 58.3%

of men and 35.3% of women decide to participate in the tournament. Although the willingness to select a tournament compensation once there is a negative externality from that decision generally reduces for all, the gender gap increases in the highest performance quartile. This result is explained by a reduction in women's willingness to participate in tournaments (only 17.7% of women compared to 58.3% of men opt into a tournament with an externality).



(a) Stage 2 (no externality)



(b) Stage 3 (externality)

Figure 3.3: Proportion of participants that choose tournament by performance quartile, by gender

These results are confirmed with a regression analysis. Table 3.3 presents the OLS regressions[15] of the decision to select a tournament compensation in stages 2 (columns (1) to (3)) and 3 (columns (4) to (6)). Controlling for the performance in the task does not significantly affect selection into tournaments with or without externality, but gender does. Women select the tournament without externality less often than men, and these differences are explained by prosociality and risk preferences. Notably, women are significantly more risk-averse than men in my sample (a Kruskal-Wallis test gives a $\chi^2 = 20.933$ and $p < 0.01$).[16] In addition, participants whose mothers' schooling level is above that of fathers', the less altruistic, and the most risk-loving, choose the tournament compensation in Stage 2 more often than others. When the tournament payment scheme involves a negative externality on others, the gender gap remains and is explained by women relatively disliking of risks. Here, older subjects and the more risk-loving tend to select the tournament alternative more often. Oppositely,

[15]Results are robust to a Probit specification.
[16]This is consistent with the finding by Dohmen et al. (2011).

prosociality and the number of male siblings favor the selection of an ex-ante egalitarian alternative. A within-subjects analysis (columns (7) to (9)) shows, in addition, that externalities reduce the frequency with which a tournament is selected, regardless of the gender of the DM. Table C.1 in the Appendix presents similar results for the top performing participants. However, the gender gap in a tournament selection in Stage 3 remains even after controlling for risk preferences.

Table 3.3: OLS regressions of decision to select a tournament compensation, with and without externality

	Stage 2			Stage 3			Stages 2 and 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.119*	-0.095	-0.040	-0.138**	-0.098*	-0.043	-0.128***	-0.097**	-0.041
	(0.067)	(0.067)	(0.072)	(0.056)	(0.054)	(0.056)	(0.049)	(0.048)	(0.048)
Correct encryptions during task	0.038*	0.035	0.032	0.028	0.023	0.020	0.033*	0.029*	0.026
	(0.022)	(0.022)	(0.021)	(0.020)	(0.018)	(0.017)	(0.018)	(0.017)	(0.016)
Age	0.032*	0.031*	0.026	0.035***	0.034**	0.028**	0.034***	0.032***	0.027**
	(0.016)	(0.016)	(0.016)	(0.013)	(0.014)	(0.014)	(0.011)	(0.011)	(0.011)
Number of female siblings	-0.028	-0.034	-0.040	-0.025	-0.036	-0.041	-0.026	-0.035	-0.040
	(0.042)	(0.042)	(0.042)	(0.033)	(0.032)	(0.032)	(0.029)	(0.029)	(0.029)
Number of male siblings	-0.028	-0.029	-0.031	-0.067**	-0.070**	-0.072**	-0.047*	-0.050**	-0.052**
	(0.037)	(0.037)	(0.036)	(0.031)	(0.029)	(0.029)	(0.025)	(0.023)	(0.023)
Economic status	0.004	0.002	-0.006	0.027	0.023	0.015	0.016	0.013	0.004
	(0.032)	(0.032)	(0.031)	(0.026)	(0.025)	(0.024)	(0.024)	(0.023)	(0.022)
Mother's schooling above father's	0.192***	0.210***	0.208***	0.018	0.048	0.046	0.105**	0.129***	0.127**
	(0.070)	(0.070)	(0.070)	(0.058)	(0.056)	(0.055)	(0.051)	(0.049)	(0.050)
Economics or finance student	0.076	0.044	0.061	0.131**	0.076	0.093	0.104**	0.060	0.077
	(0.070)	(0.071)	(0.071)	(0.061)	(0.059)	(0.059)	(0.052)	(0.051)	(0.050)
Preference for equal payment conditions		-0.158**	-0.124*		-0.270***	-0.237***		-0.214***	-0.181***
		(0.068)	(0.069)		(0.051)	(0.051)		(0.046)	(0.046)
Willingness to take risks			0.074**			0.073***			0.073***
			(0.031)			(0.024)			(0.021)
Externality							-0.118***	-0.118***	-0.118***
							(0.036)	(0.036)	(0.037)
Constant	-0.655	-0.543	-0.629	-0.753**	-0.561	-0.645*	-0.645**	-0.493	-0.578**
	(0.434)	(0.440)	(0.436)	(0.355)	(0.377)	(0.364)	(0.297)	(0.305)	(0.290)
Observations	203	203	203	203	203	203	406	406	406
Clusters	-	-	-	-	-	-	203	203	203

Notes: OLS regressions. Dependent variable is the preference for tournament payment. Columns (1) to (3) include the decisions without externality (Stage 2); Columns (4) to (6) include the decisions with externality (Stage 3). Columns (7) to (9) include the decisions in both stages, and Externality is a dummy variable that takes the value of 1 for Stage 3. Robust standard errors in parenthesis in Columns (1) to (6); Standard errors clustered at the individual level (clusters=203) in Columns (7) to (8). *** p<0.01, ** p<0.05, * p<0.1.

Result 4: At any performance level, women are less likely than men to select the tournament compensation, with or without externality. This is explained by risk preferences.

Result 5: Externalities reduce the frequency of selection of a tournament compensation.

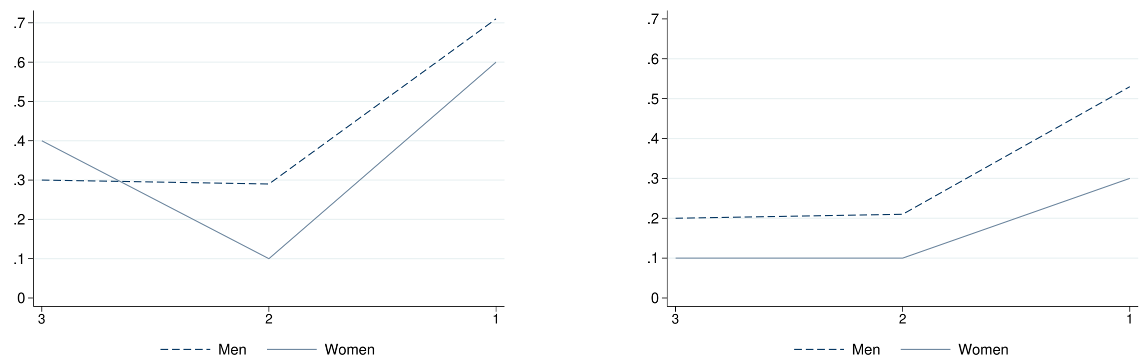
Next, I study how subjects form their beliefs about their relative position in their group and the session and whether tournament selection depends on their perceived ranks. 24.27% of women and 34% of men think they are the top performers of their respective groups, even when they do not have any information on the group composition. I measure overconfidence as the distance between the belief and the actual position in the group and find that 25.24% of women and 29% of men are overconfident. Once I consider the subjects' relative position in the session, I find that 37% of women and 48% of men are overconfident regarding their position. Table 3.4 presents an ordered probit of the beliefs of the relative position within the group and the session. Note that a position of 1 indicates that a subject believes she is the best in her group (session). Subjects with relatively higher performance believe they have a better position in the group and the session. Women are less optimistic about their relative performance in the session but not in the group. The beliefs of relative position do not depend on the percentage of women in the session.

Table 3.4: Ordered probit of belief of relative position

	Position in Group	Position in Session	
	(1)	(2)	(3)
Correct encryptions during task	-0.453***	-0.317***	-0.318***
	(0.066)	(0.053)	(0.053)
Female	0.143	0.257*	0.283*
	(0.174)	(0.146)	(0.151)
% of women in session			-0.484
			(0.612)
Observations	203	203	203
Controls	Yes	Yes	Yes

Notes: Ordered Probit regressions. Dependent variables are the beliefs on relative position in the group (Column (1)) and in the session (columns (2) and (3)). Controls: age, number of female and male siblings, economic status, mother's schooling level above that of father's, and economics or finance student. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Figure 3.4 shows the proportion of women and men that submit their performance to a tournament in stages 2 (3.4a) and 3 (3.4b). In both cases, women and men who believe they are the first on their group are more likely to decide to participate in the tournament. There is still a gender gap in willingness to submit their performance to a tournament, both with and without externality. A regression analysis available in Table C.2 in the Appendix shows that the gender gap is not statistically significant in these cases. However, the decision to select the agentic compensation reduces for all in the presence of a negative externality, and the gender gap is greater for those who believe that they are in the first position.



(a) Stage 2 (no externality)

(b) Stage 3 (externality)

Figure 3.4: Proportion of participants that choose tournament by belief of position in group, by gender

Table 3.5 presents the OLS estimations results for the decision to select a tournament compensation depending on the beliefs of relative performance. Columns (1) to (4) present the results for Stage 2; Columns (5) to (8) show the results for Stage 3. Columns (9) and (12) present the within-subject analysis by pooling results.

Table 3.5: OLS regressions of decision to select a tournament compensation, with and without externality

	Stage 2				Stage 3				Stages 2 and 3			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.103*	-0.079	-0.055	-0.043	-0.129**	-0.089*	-0.065	-0.045	-0.116**	-0.084*	-0.060	-0.044
	(0.062)	(0.064)	(0.067)	(0.068)	(0.055)	(0.053)	(0.058)	(0.055)	(0.046)	(0.045)	(0.045)	(0.045)
Belief of relative position in Group 1	-0.288***	-0.289***	-0.263***	-0.268***	-0.162***	-0.164***	-0.128**	-0.138***	-0.225***	-0.226***	-0.195***	-0.203***
	(0.069)	(0.068)	(0.070)	(0.069)	(0.058)	(0.052)	(0.058)	(0.053)	(0.049)	(0.046)	(0.049)	(0.046)
Correct encryptions during task	-0.016	-0.019	-0.014	-0.017	-0.002	-0.008	-0.000	-0.005	-0.009	-0.013	-0.007	-0.011
	(0.026)	(0.026)	(0.025)	(0.025)	(0.022)	(0.020)	(0.021)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)
Age	0.026*	0.025*	0.022	0.022	0.032**	0.030**	0.026**	0.026*	0.029***	0.028***	0.024**	0.024**
	(0.014)	(0.014)	(0.015)	(0.014)	(0.013)	(0.015)	(0.013)	(0.014)	(0.010)	(0.010)	(0.010)	(0.010)
Number of female siblings	-0.019	-0.026	-0.025	-0.030	-0.020	-0.031	-0.028	-0.036	-0.019	-0.028	-0.027	-0.033
	(0.037)	(0.038)	(0.037)	(0.038)	(0.033)	(0.031)	(0.033)	(0.031)	(0.026)	(0.026)	(0.026)	(0.026)
Number of male siblings	-0.026	-0.027	-0.028	-0.029	-0.066**	-0.069**	-0.069**	-0.071**	-0.046*	-0.048**	-0.048**	-0.050**
	(0.034)	(0.034)	(0.034)	(0.033)	(0.031)	(0.029)	(0.031)	(0.029)	(0.023)	(0.021)	(0.023)	(0.021)
Economic status	0.001	-0.001	-0.006	-0.006	0.026	0.022	0.016	0.015	0.014	0.011	0.005	0.005
	(0.028)	(0.028)	(0.028)	(0.028)	(0.024)	(0.023)	(0.023)	(0.023)	(0.021)	(0.020)	(0.020)	(0.020)
Mother's schooling above father's	0.214***	0.232***	0.213***	0.229***	0.030	0.060	0.030	0.057	0.122**	0.146***	0.122**	0.143***
	(0.066)	(0.066)	(0.067)	(0.067)	(0.058)	(0.055)	(0.057)	(0.054)	(0.048)	(0.047)	(0.048)	(0.047)
Economics or finance student	0.092	0.059	0.099	0.070	0.140**	0.084	0.149**	0.097*	0.116**	0.072	0.124**	0.083*
	(0.066)	(0.066)	(0.066)	(0.067)	(0.060)	(0.058)	(0.060)	(0.058)	(0.049)	(0.047)	(0.048)	(0.047)
Preference for equal payment conditions		-0.160**		-0.138**		-0.271***		-0.244***		-0.216***		-0.191***
		(0.065)		(0.067)		(0.050)		(0.050)		(0.044)		(0.045)
Willingness to take risks			0.060*	0.049			0.080***	0.060**			0.070***	0.054**
			(0.031)	(0.032)			(0.025)	(0.024)			(0.022)	(0.021)
Externality									-0.118***	-0.118***	-0.118***	-0.118***
									(0.036)	(0.037)	(0.037)	(0.037)
Constant	0.420	0.537	0.274	0.403	-0.149	0.050	-0.342	-0.114	0.194	0.353	0.025	0.204
	(0.465)	(0.462)	(0.475)	(0.472)	(0.426)	(0.446)	(0.420)	(0.437)	(0.338)	(0.343)	(0.336)	(0.336)
Observations	203	203	203	203	203	203	203	203	406	406	406	406
Clusters	-	-	-	-	-	-	-	-	203	203	203	203

Notes: OLS regressions. Dependent variable is the preference for tournament payment. Columns (1) to (4) include the decisions without externality (Stage 2); Columns (5) to (8) include the decisions with externality (Stage 3). Columns (9) to (12) include the decisions in both stages, and Externality is a dummy variable that takes the value of 1 for Stage 3. Robust standard errors in parenthesis in Columns (1) to (8); Standard errors clustered at the individual level (clusters=203) in Columns (9) and (12). *** p<0.01, ** p<0.05, * p<0.1.

Overall, more men than women select a competitive compensation, even when it does not require them to perform under that payment scheme. Prosocialness and risk preferences explain the gender gap in Stage 2, while risk preferences alone explain the gender gap in selecting a tournament compensation in Stage 3. Furthermore, unlike actual performance, the beliefs of the relative position in the group affect compensation choices. In particular, the highest the guessed ranking, the least likely it is for subjects to select a tournament in stages 2 and 3. However, beliefs do not eliminate the gender gap in choosing an agentic over a communal norm. Having a mother whose education level is above that of the father, self-serving preferences, and risk-loving attitudes increase the selection of a tournament compensation when there are no negative externalities. When choosing a tournament can potentially affect others, older participants, economics and finance students, and those with fewer male siblings also select the agentic norm more frequently than others. The within-subject analysis shows that externalities reduce the willingness to select a tournament compensation, even after controlling for altruism and risk preferences.

Result 6: Beliefs of relative position in the group affect selection of a tournament compensation, but do not eliminate the gender gap.

3.3.2 Group decisions

No information on group composition

I classify participants according to the decisions in stages 3 and 4 and find that 54.4% of women and 36% of men always chose Alternative 1, that 7% of women and 25% of men always choose Alternative 2, and that 30.1% of women and 32% of men select the competitive compensation only when they are part of a group ($\chi^2 = 14.70$ and $p < 0.01$). Table 3.6 shows the results for the OLS estimations of the decisions in Stage 4 (columns (1) to (4)) and for the within-subject decisions in stages 3 and 4, where individual tournament participation becomes a group's tournament participation (columns (5) to (7)). I see that after controlling for ability[17] and the belief on the relative position in the session,[18] women choose a tournament compensation as part of a group less than men. This gender gap remains significant after controlling for prosociality but disappears once controlling for risk aversion. In addition, as subjects become more risk-loving, the decision to partake in a tournament increases. Oppositely, the more altruistic subjects are less likely to select a competitive compensation. A within-subject analysis shows that group participation increases the selection of a tournament compensation scheme compared to when decisions happen individually. Contrary to Healy and Pate (2011), I do not find that women compete more in groups when there is no information on group composition.

[17]Women and men who choose Alternative 1 are not statistically different in terms of RET performance from those who select Alternative 2 ($p = 0.643$ for women and $p = 0.320$ for men).

[18]I did not elicit beliefs of the relative position within the anonymous group or between anonymous groups in a supergroup.

Table 3.6: OLS regressions of decision to select a tournament compensation with externalities, in groups

	Stage 4				Stages 3 and 4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.169**	-0.124*	-0.114	-0.069	-0.130**	-0.087	-0.038
	(0.073)	(0.071)	(0.072)	(0.074)	(0.056)	(0.054)	(0.056)
Belief of relative position in session	-0.009	-0.010	-0.008	-0.003	-0.013	-0.013	-0.007
	(0.015)	(0.014)	(0.014)	(0.014)	(0.010)	(0.009)	(0.009)
Correct encryptions during task	0.012	0.006	0.005	0.007	0.013	0.008	0.009
	(0.027)	(0.026)	(0.026)	(0.026)	(0.018)	(0.016)	(0.015)
Preference for equal payment conditions		-0.303***	-0.274***	-0.253***		-0.287***	-0.255***
		(0.071)	(0.075)	(0.077)		(0.045)	(0.046)
Decision to participate in tournament in stage 3			0.108	0.071			
			(0.086)	(0.087)			
Willingness to take risks				0.069**			0.071***
				(0.032)			(0.021)
Group					0.250***	0.250***	0.250***
					(0.058)	(0.058)	(0.058)
Female × Group					-0.036	-0.036	-0.036
					(0.083)	(0.083)	(0.083)
Constant	0.417	0.638	0.668	0.480	-0.151	0.058	-0.124
	(0.530)	(0.517)	(0.515)	(0.526)	(0.346)	(0.353)	(0.351)
Observations	203	203	203	203	406	406	406
Clusters	-	-	-	-	203	203	203
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions. Dependent variable is the preference for tournament payment in groups with no information on gender composition. Columns (1) to (4) include the individual preferences for group's tournament with externality (Stage 4); Columns (5) to (7) include individual and group decisions with externality (Stages 3 and 4). Group tournament is a dummy variable that takes the value of 1 for Stage 4. Controls: age, number of female and male siblings, economic status, mother's schooling level above that of father's, and economics or finance student. Robust standard errors in parenthesis in Columns (1) to (5); Standard errors clustered at the individual level (clusters=203) in Columns (5) to (7). *** p<0.01, ** p<0.05, * p<0.1.

Result 7: As part of an anonymous group, women are also less likely to select a tournament compensation than men. This is explained by individual preferences for tournament participation and willingness to take risks.

Result 8: Group participation increases the selection of a tournament compensation, regardless of gender.

Table 3.7 shows the results for the OLS regressions for the decision to select a tournament compensation of subjects that, as individuals, chose Alternative 1. The dependent variable takes zero value for participants that choose Alternative 1 in stages 3 and 4. I do not observe a gender gap in changing from an egalitarian to an agentic distribution norm. However, the more risk-loving and the less prosocial tend to change more frequently from an egalitarian norm as individuals to a tournament compensation as part of a group.

Table 3.7: OLS regressions of decision to select a tournament compensation in groups when individually selecting not to

	Changes between stages 3 and 4		
	(1)	(2)	(3)
Female	-0.109	-0.077	-0.025
	(0.083)	(0.080)	(0.083)
Belief of relative position in session	0.002	-0.004	0.002
	(0.016)	(0.016)	(0.016)
Correct encryptions during task	0.026	0.018	0.018
	(0.032)	(0.030)	(0.030)
Preference for equal payment conditions		-0.281***	-0.259***
		(0.079)	(0.080)
Willingness to take risks			0.078**
			(0.037)
Constant	0.167	0.492	0.285
	(0.647)	(0.640)	(0.644)
Observations	155	155	155
Controls	Yes	Yes	Yes

Notes: OLS regressions. Dependent variable is the preference for tournament payment as part of a group when individual preference was for Alternative 1. Controls: age, number of female and male siblings, economic status, mother's schooling level above that of father's, and economics or finance student. Robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.

Result 9: There is no gender gap in changing from an egalitarian payment scheme in Stage 3 to a tournament compensation in Stage 4. More risk loving and less prosocial subjects are more likely to make this change.

With information on group composition

After classifying subjects according to the decisions in stages 4 and 5, I find that 52.4% of women and 33% of men never select the tournament option when they are part of a group, while 21.4% of women and 47% of men always do.[19] Accounting for the group's gender composition, I see that 55% of men in male-majority groups (FMM) and 59.46% of men in female-majority groups (FFM) choose to select a tournament compensation ($\chi^2 = 0.185$ and $p = 0.667$). In turn, 41.38% of women in male-majority groups and 28.38% of women in female-majority groups choose a tournament compensation

[19] Considering the changes in decisions from Stage 3 to Stage 5, 64% of women and 40% of men do not select a tournament compensation in both stages.

($\chi^2 = 1.617$ and $p = 0.203$). Although there are no statistically significant differences in selection in a tournament by gender of the decision maker, when pooled, male-majority groups favor an agentic behavior over a communal one. 50.56% of subjects in male-majority groups and 38.74% of subjects in female-majority groups choose competition ($\chi^2 = 2.802$ and $p = 0.094$).

Keeping only the subjects who change from one alternative to the other between stages 4 and 5 (N=47), 60% of men in male-majority groups change towards tournament compensation. In contrast, 60% of men in female-majority groups change towards an egalitarian compensation rule ($\chi^2 = 0.080$ and $p = 0.371$). Similarly, 50% of women in a male-majority group change to Alternative 2, while 61.9% of women in female-majority groups change to Alternative 1 ($\chi^2 = 0.274$ and $p = 0.601$). When pooled, 56% of those in male-majority groups opt in a tournament, while 61% of those in female-majority groups opt out of a tournament ($\chi^2 = 1.314$ and $p = 0.252$).

Women who decide to participate in the tournament as part of a group with information are not statistically different in terms of performance in the RET from those who decide not to. On average, women that choose Alternative 1 in stage 5 have an average performance of 8.14, while women that choose Alternative 2 have an average performance of 8.30 ($t = -0.477$ and $p = 0.635$). Oppositely, men who select Alternative 2 in Stage 5 have a relatively better performance than those who do not (8.67 vs. 8.12, $t = -1.822$ and $p = 0.074$). Figure 3.5 shows the proportion of women and men who select the tournament in Stage 5 by performance quartile. Men are more likely than women to select a tournament compensation for every performance level. The gender gap widens for the top performers (quartiles 3 and 4), with 60% or more men selecting this option and 30% of women choosing a tournament compensation that generates a negative externality. These results are confirmed with a regression analysis (see Tables C.3 and C.4 in the Appendix) and the channel is risk preferences.

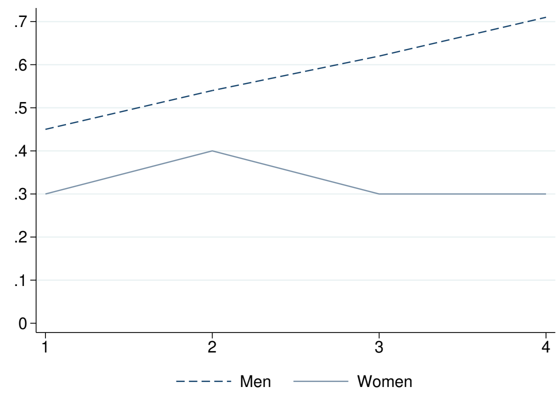
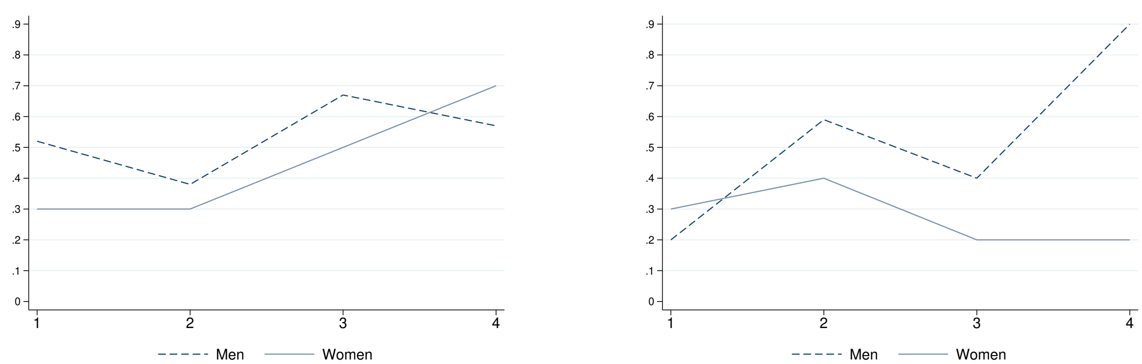


Figure 3.5: Proportion of Participants that Choose Tournament by Performance Quartile, by Gender

To study how gender composition favors changes toward an egalitarian rule, I begin by breaking Figure 3.5 into majority-male and majority-female groups. Figure 3.6 shows the proportion of participants selecting a tournament divided by group composition. Here, it is apparent that the gender gap in preference for competitive compensation disappears in male-majority groups. However, high achieving women appear to be 10pp more likely than high achieving men to select in tournaments. In contrast, the gender gap in tournament selection remains in female-majority groups, except for the lowest performance quartile, and is notably higher for the top performers.



(a) Male-majority Groups

(b) Female-majority Groups

Figure 3.6: Proportion of Participants that Choose Tournament by Performance Quartile, by Gender and Group Composition

Regarding beliefs, 46% of men and 20.4% of women believe they are the top performers within their groups. Here, 33% of men and 24.3% of women believe their position is above their actual rank. Women, on average, are underconfident in male-majority groups ($t = -1.556$ and $pr(T < t) = 0.06$) while men are equally confident in female and male majority groups ($t = -0.477$ and $pr(|T| > |t|) = 0.635$). Table 3.8 presents the results for the OLS regressions of the beliefs on the relative position within the group and the position of one's group in the supergroup, and concerning an anonymous randomly selected group in the session. Subjects with relatively higher performance believe they have a better position in the group. A higher number of correct encryptions also increases the probability of believing that one's group is better than another randomly selected group within the session. However, it does not affect the belief of the relative performance of one's group versus the paired group. In addition, women are less optimistic about their relative performance in the group, more so in female-majority groups, than men in the same group composition.

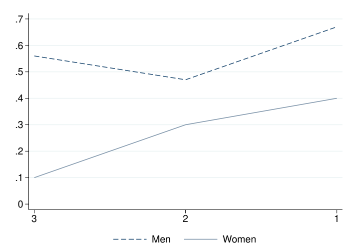
Table 3.8: OLS regressions of beliefs of individual and group's relative position

	Position in group (1)	Own group is better (2)	Own group is better than random (3)
Correct encryptions during task	-0.170***	0.010	0.038*
	(0.024)	(0.016)	(0.021)
Female	0.362***	-0.093	0.002
	(0.133)	(0.080)	(0.106)
Female majority	-0.167	0.028	0.045
	(0.113)	(0.046)	(0.093)
Female × Female majority	-0.009	0.031	-0.136
	(0.171)	(0.093)	(0.141)
Constant	3.634***	0.713**	0.183
	(0.531)	(0.306)	(0.382)
Observations	203	203	203
Controls	Yes	Yes	Yes
Tests of joint coefficients:			
Additional effect of women in FFM	0.353	-0.062	-0.135
p-val	< 0.01	0.175	0.153

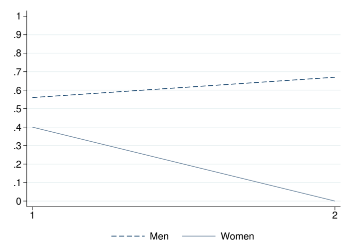
Notes: the dependent variables are the belief on relative position in the group (Column (1)) and the beliefs that own group is better than other group from the same supergroup (Column (2)) or from other randomly determined group in the session (Column (3)). Joint corresponds to the additional effect of women in female majority groups. Controls: age, number of female and male siblings, economic status, mother's schooling level above that of father's, economics or finance student, and all-men groups. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Result 10: Women are underconfident in male-majority groups while men are equally confident in both group configurations.

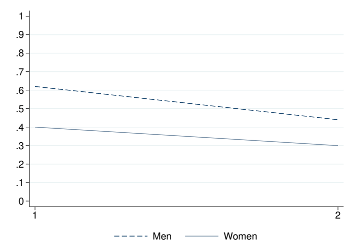
Figure 3.7 shows the proportion of women and men that submit their performance to a tournament in Stage 5 depending on believed rank in group (3.7a), and believed position of group with respect to other group in supergroup (3.7b) and to other randomly selected group in the session (3.7c). Women and men who believe they rank first in their group are more likely to decide to participate in the tournament. Similarly, the believed position of own group concerning the other group generally affects the selection into a competitive compensation. A notable case is that of men who believe their group ranks second in the supergroup and still select the tournament (67% of men vs 0% of women).



(a) Position in group



(b) Position of group wrt other group



(c) Position of group wrt random group

Figure 3.7: Proportion of Participants that Choose Tournament in Stage 5 by Beliefs, by Gender

Table 3.9 presents the OLS regression results for selecting a tournament compensation when there is information on group composition. As suggested from the figures above, female-majority groups favor a more egalitarian norm than male-majority groups, even after controlling for performance and beliefs on the relative performance of the group concerning the other group from the same supergroup. This is explained by sociodemographic characteristics (age, number of female and male siblings, economic status, and mother's schooling level above that of father's). Altruism, risk, and preferences for tournament compensation further reduce the gap between male and female majority groups. Once controlling for female participants, the marginal effect is smaller. In addition, Table C.5 in the Appendix shows the OLS regressions results after pooling stages 3 and 5 and 4 and 5. The increase in selection into a tournament seems to be a consequence of group competition. Adding information on gender composition does not appear to affect the choice of a compensation scheme.

Table 3.9: OLS regressions of decision to select a tournament compensation with information of group composition

	Stage 5							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female majority	-0.139**	-0.113	-0.127*	-0.094	-0.054	-0.055	-0.034	-0.020
	(0.069)	(0.070)	(0.070)	(0.072)	(0.063)	(0.054)	(0.054)	(0.054)
Correct encryptions during task	0.040**	0.046**	0.036*	0.040**	0.032*	0.024	0.020	0.020
	(0.020)	(0.020)	(0.021)	(0.020)	(0.019)	(0.017)	(0.016)	(0.016)
Belief group is better that other group in super-group	0.240**	0.284**	0.245**	0.290***	0.304***	0.346***	0.327***	0.318***
	(0.105)	(0.111)	(0.105)	(0.109)	(0.106)	(0.101)	(0.100)	(0.099)
Age		0.003		0.008	0.005	0.002	-0.004	-0.003
		(0.016)		(0.017)	(0.017)	(0.016)	(0.016)	(0.016)
Number of female siblings		0.053		0.064	0.051	0.071**	0.066**	0.068**
		(0.040)		(0.042)	(0.036)	(0.033)	(0.033)	(0.033)
Number of male siblings		0.037		0.045	0.038	0.029	0.025	0.023
		(0.043)		(0.044)	(0.040)	(0.034)	(0.035)	(0.034)
Economic status		-0.080**		-0.083**	-0.092***	-0.084***	-0.095***	-0.095***
		(0.032)		(0.033)	(0.029)	(0.026)	(0.025)	(0.026)
Mother's schooling above father's		-0.081		-0.078	-0.033	-0.030	-0.034	-0.036
		(0.071)		(0.070)	(0.064)	(0.055)	(0.054)	(0.053)
Economics or finance student			0.066	0.117	0.024	-0.008	0.004	-0.003
			(0.074)	(0.076)	(0.072)	(0.059)	(0.058)	(0.059)
All men groups			0.210	0.204	0.255	0.203	0.238	0.222
			(0.184)	(0.227)	(0.320)	(0.167)	(0.214)	(0.216)
Preference for equal payment conditions					-0.406***	-0.260***	-0.232***	-0.230***
					(0.066)	(0.067)	(0.065)	(0.065)
Decision to participate in tournament in stage 4						0.450***	0.410***	0.406***
						(0.064)	(0.068)	(0.069)
Willingness to take risks							0.076***	0.071***
							(0.023)	(0.025)
Female								-0.046
								(0.060)
Constant	-0.032	0.034	-0.043	-0.079	0.242	0.044	0.027	0.066
	(0.184)	(0.433)	(0.188)	(0.434)	(0.454)	(0.425)	(0.409)	(0.412)
Observations	203	203	203	203	203	203	203	203

Notes: OLS regressions. Dependent variable is the preference for tournament payment in groups with information of gender composition. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Result 11: Female majority groups favor an egalitarian norm over a tournament compensation. This is explained by the gender of the decision-maker, altruism, risk and competition preferences, and sociodemographic characteristics.

To further explore the differences in the behavior of men and women in male and female-majority groups, Table 3.10 presents the OLS regressions, including an interaction term for both variables. Tests of joint coefficients confirm that the probability of selecting a tournament compensation scheme decreases for women in female-majority groups ($p < 0.01$). This is robust after controlling for performance and beliefs on relative performance between both groups of the super-group, altruism, and preferences on group tournament compensation, and marginally robust to risk aversion. In addition, after controlling for altruism, women in male-majority groups are less likely to select the agentic norm than men in the same groups. Overall, the probability of selecting an egalitarian compensation scheme is largest for women in both group configurations. No particular group configuration favors men moving to a communal norm. Only the most altruistic men in both female and male-majority groups select the egalitarian alternative over the agentic norm.

Table 3.10: OLS regressions of decision to select a tournament compensation with information of group composition

	Stage 5									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female majority	0.061	0.039	0.015	-0.006	0.007	-0.015	0.040	0.020	-0.022	-0.041
	(0.100)	(0.100)	(0.086)	(0.086)	(0.086)	(0.088)	(0.095)	(0.096)	(0.076)	(0.078)
Female	-0.174	-0.149	-0.196**	-0.169*	-0.121	-0.089	-0.092	-0.070	-0.099	-0.069
	(0.116)	(0.114)	(0.097)	(0.095)	(0.092)	(0.085)	(0.109)	(0.108)	(0.081)	(0.076)
Female majority × Female	-0.126	-0.120	0.011	0.013	-0.051	-0.049	-0.075	-0.073	0.045	0.044
	(0.153)	(0.149)	(0.131)	(0.129)	(0.133)	(0.127)	(0.148)	(0.144)	(0.119)	(0.114)
Age	0.005	0.006	0.004	0.005	0.002	0.002	-0.005	-0.003	-0.004	-0.003
	(0.017)	(0.017)	(0.018)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Number of female siblings	0.073*	0.074*	0.056	0.057	0.084**	0.086**	0.061	0.062	0.064*	0.067**
	(0.042)	(0.041)	(0.037)	(0.036)	(0.035)	(0.034)	(0.039)	(0.038)	(0.034)	(0.033)
Number of male siblings	0.031	0.037	0.026	0.032	0.022	0.028	0.025	0.031	0.017	0.023
	(0.044)	(0.044)	(0.042)	(0.040)	(0.037)	(0.037)	(0.044)	(0.043)	(0.036)	(0.034)
Economic status	-0.073**	-0.084**	-0.084***	-0.096***	-0.066**	-0.079***	-0.092***	-0.102***	-0.085***	-0.096***
	(0.033)	(0.033)	(0.030)	(0.030)	(0.028)	(0.028)	(0.031)	(0.031)	(0.027)	(0.026)
Mother's schooling above father's	-0.091	-0.083	-0.049	-0.041	-0.067	-0.059	-0.086	-0.079	-0.044	-0.036
	(0.069)	(0.068)	(0.063)	(0.063)	(0.059)	(0.057)	(0.066)	(0.066)	(0.055)	(0.054)
Economics or finance student	0.093	0.075	0.018	0.003	0.038	0.022	0.110	0.094	0.013	-0.001
	(0.076)	(0.077)	(0.072)	(0.072)	(0.062)	(0.061)	(0.076)	(0.076)	(0.061)	(0.060)
All men groups	0.081	0.163	0.114	0.201	0.042	0.141	0.168	0.246	0.119	0.214
	(0.308)	(0.243)	(0.398)	(0.328)	(0.176)	(0.107)	(0.378)	(0.311)	(0.300)	(0.219)
Belief group is better that other group in super-group		0.249**		0.271***		0.320***		0.247**		0.317***
		(0.102)		(0.100)		(0.095)		(0.110)		(0.099)
Correct encryptions during task		0.035*		0.029		0.026		0.029		0.020
		(0.019)		(0.018)		(0.017)		(0.019)		(0.017)
Preference for equal payment conditions			-0.386***	-0.386***					-0.233***	-0.234***
			(0.067)	(0.066)					(0.068)	(0.065)
Decision to participate in tournament in stage 4					0.496***	0.506***			0.394***	0.406***
					(0.063)	(0.060)			(0.073)	(0.069)
Willingness to take risks							0.131***	0.128***	0.075***	0.071***
							(0.031)	(0.030)	(0.027)	(0.025)
Constant	0.599	0.094	0.836**	0.367	0.404	-0.078	0.452	0.005	0.503	0.072
	(0.406)	(0.438)	(0.418)	(0.454)	(0.369)	(0.399)	(0.390)	(0.427)	(0.375)	(0.411)
Observations	203	203	203	203	203	203	203	203	203	203
Tests of joint coefficients:										
Men in FFM	0.660	0.132	0.851	0.362	0.410	-0.093	0.491	0.025	0.481	0.031
p-val	0.104	0.763	$p < 0.05$	0.425	0.274	0.821	0.206	0.953	0.203	0.941
Women in FFM	0.360	-0.138	0.665	0.205	0.238	-0.231	0.324	-0.117	0.427	0.006
p-val	0.373	0.750	0.111	0.647	0.515	0.561	0.398	0.779	0.250	0.989
Women in FMM	0.425	-0.056	0.640	0.198	0.282	-0.167	0.359	-0.065	0.404	0.003
p-val	0.300	0.897	0.133	0.664	0.462	0.678	0.356	0.875	0.294	0.994
Additional effect of Women in FFM	-0.300	-0.270	-0.185	-0.156	-0.172	-0.139	-0.167	-0.143	-0.053	-0.025
p-val	$p < 0.01$	$p < 0.01$	$p < 0.05$	$p < 0.1$	$p < 0.1$	0.140	0.100	0.156	0.554	0.776

Notes: OLS regressions. Dependent variable is the preference for tournament payment in groups with information of gender composition. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Result 12: No particular group composition favors men moving towards a more egalitarian norm.

Next, I study the decisions of subjects who change from Alternative 1 in Stage 3 to Alternative 2 in Stage 5. I create a variable that takes the value of 1 for those who, during Stage 3 select the egalitarian compensation scheme and, during Stage 5, select the tournament alternative. The variable takes the value of 0 for who that always select the egalitarian alternative. Table 3.11 shows the OLS estimations results for those who change decisions between these two stages.

Table 3.11: OLS regressions of decision to change from an egalitarian alternative to tournament compensation

	Stage 5									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female majority	0.079	0.060	0.019	0.001	0.036	0.011	0.071	0.055	0.003	-0.018
	(0.113)	(0.112)	(0.101)	(0.101)	(0.100)	(0.102)	(0.112)	(0.111)	(0.092)	(0.094)
Female	-0.082	-0.081	-0.119	-0.118	-0.054	-0.043	-0.037	-0.035	-0.056	-0.046
	(0.124)	(0.123)	(0.106)	(0.106)	(0.105)	(0.099)	(0.119)	(0.119)	(0.094)	(0.092)
Female majority × Female	-0.179	-0.159	-0.058	-0.040	-0.132	-0.106	-0.144	-0.126	-0.046	-0.027
	(0.160)	(0.157)	(0.137)	(0.136)	(0.145)	(0.141)	(0.159)	(0.156)	(0.129)	(0.126)
Age	-0.019	-0.020	-0.020	-0.021	-0.022	-0.024*	-0.024	-0.026	-0.025*	-0.028*
	(0.018)	(0.018)	(0.017)	(0.016)	(0.015)	(0.014)	(0.017)	(0.017)	(0.014)	(0.014)
Number of female siblings	0.098**	0.103**	0.079**	0.083**	0.109***	0.117***	0.089**	0.093**	0.090**	0.098***
	(0.042)	(0.041)	(0.039)	(0.039)	(0.036)	(0.035)	(0.040)	(0.040)	(0.036)	(0.035)
Number of male siblings	0.100**	0.095**	0.097**	0.092**	0.083**	0.074*	0.098**	0.093**	0.083**	0.074**
	(0.046)	(0.047)	(0.044)	(0.044)	(0.039)	(0.040)	(0.044)	(0.045)	(0.036)	(0.037)
Economic status	-0.105***	-0.111***	-0.111***	-0.116***	-0.097***	-0.105***	-0.115***	-0.120***	-0.107***	-0.114***
	(0.032)	(0.032)	(0.030)	(0.030)	(0.028)	(0.028)	(0.032)	(0.032)	(0.027)	(0.027)
Mother's schooling above father's	-0.040	-0.044	-0.004	-0.008	-0.016	-0.022	-0.037	-0.041	0.003	-0.003
	(0.075)	(0.074)	(0.073)	(0.072)	(0.065)	(0.063)	(0.073)	(0.072)	(0.064)	(0.062)
Economics or finance student	0.048	0.034	-0.001	-0.013	0.040	0.024	0.068	0.055	0.024	0.009
	(0.081)	(0.082)	(0.077)	(0.079)	(0.068)	(0.066)	(0.080)	(0.081)	(0.066)	(0.065)
All men groups	-0.662***	-0.449**	-0.825***	-0.614***	-0.399***	-0.051	-0.741***	-0.529***	-0.590***	-0.255
	(0.173)	(0.201)	(0.160)	(0.186)	(0.153)	(0.183)	(0.167)	(0.202)	(0.154)	(0.192)
Belief group is better that other group in super-group		0.195*		0.194*		0.323***		0.198*		0.303***
		(0.106)		(0.100)		(0.109)		(0.115)		(0.113)
Correct encryptions during task		0.021		0.019		0.012		0.015		0.008
		(0.018)		(0.018)		(0.016)		(0.019)		(0.017)
Preference for equal payment conditions			-0.293***	-0.291***					-0.178**	-0.169**
			(0.069)	(0.069)					(0.071)	(0.071)
Decision to participate in tournament in stage 4					0.429***	0.456***			0.354***	0.383***
					(0.069)	(0.066)			(0.081)	(0.079)
Willingness to take risks							0.094***	0.093***	0.054*	0.051*
							(0.035)	(0.035)	(0.031)	(0.030)
Constant	0.994**	0.689	1.213***	0.930**	0.833**	0.525	0.856**	0.605	0.916***	0.645*
	(0.413)	(0.440)	(0.400)	(0.428)	(0.341)	(0.369)	(0.413)	(0.444)	(0.347)	(0.378)
Observations	155	155	155	155	155	155	155	155	155	155
Tests of joint coefficients:										
Men in FFM	1.073	0.749	1.232	0.932	0.869	0.536	0.927	0.660	0.918	0.627
p-val	$p < 0.05$	0.100	$p < 0.01$	$p < 0.05$	$p < 0.05$	0.173	$p < 0.05$	0.145	$p < 0.05$	0.110
Women in FFM	0.812	0.509	1.054	0.774	0.684	0.387	0.746	0.498	0.816	0.554
p-val	$p < 0.1$	0.249	$p < 0.01$	$p < 0.1$	$p < 0.05$	0.301	$p < 0.1$	0.257	$p < 0.05$	0.139
Women in FMM	0.912	0.608	1.093	0.813	0.779	0.481	0.818	0.570	0.859	0.599
p-val	$p < 0.05$	0.164	$p < 0.01$	$p < 0.1$	$p < 0.05$	0.197	$p < 0.05$	0.188	$p < 0.05$	0.113
Additional effect of Women in FFM	-0.261	-0.240	-0.178	-0.158	-0.185	-0.149	-0.181	-0.162	-0.102	-0.072
p-val	$p < 0.01$	$p < 0.05$	$p < 0.1$	$p < 0.1$	$p < 0.1$	0.129	$p < 0.1$	0.126	0.277	0.435

Notes: OLS regressions. Dependent variable is the preference for tournament payment in groups with information of gender composition. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Using joint coefficients tests I find that women in female-majority groups are less likely to change from Alternative 1 in Stage 3 to Alternative 2 in Stage 3 compared to men in groups with the same gender configuration. A higher number of siblings and lower economic status favor changes from the communal to the agentic compensation. Notably, all-men groups are less likely to favor changes from a a communal to an agentic norm.

Overall, group composition appears insufficient for creating unconditional behavioral changes towards a more communal norm from non-altruistic individuals. Now I move to study the role of group leaders in creating normative changes.

3.3.3 The role of leaders

Stage 6 incorporates the role of leaders in a group. Sixty-eight participants (34 in each treatment) decide the payment rule for both groups in the supergroup. Half of the leaders are women, and half are men. 41.18% of female leaders and 76.47% of male leaders select a competitive compensation ($\chi^2 = 8.743$ and $p < 0.01$). This statistical difference holds when I consider both treatments separately ($\chi^2 = 4.250$ and $p < 0.05$ for the random treatment and $\chi^2 = 4.636$ and $p < 0.05$ for the merit treatment) (Table C.6 in the Appendix presents the summary statistics by treatment). Table 3.12 shows the OLS regressions for leader's selection into a tournament compensation scheme. Female leaders are more likely than male leaders to select the egalitarian alternative, independent of treatment. This gender gap remains significant even after controlling for the leader's performance, the beliefs of the group's relative position within the supergroup, altruism, preferences for competition, and risk.

The number of times the leader believes her choice of Alternative 2 will be accepted by other group members (0, 1, or 2) and age positively affect the selection of a competitive compensation scheme. In contrast, the more altruistic participants, those with more female siblings, and those in all-men groups are more likely to select the egalitarian alternative as leaders. However, none of these factors individually eliminates the gender gap in terms of female leaders' typically selecting the more communal norm than men. In the merit treatment, women, more than men decide against the tournament compensation.

Result 13: Female leaders are more likely to select the egalitarian alternative than male ones, regardless of treatment.

Table 3.12: OLS regressions of decision of leader to select a tournament compensation (treatment)

	Stage 6					
	(1)	(2)	(3)	(4)	(5)	(6)
Female leader	-0.340***	-0.441**	-0.396**	-0.391***	-0.391***	-0.365***
	(0.126)	(0.176)	(0.150)	(0.122)	(0.124)	(0.125)
Merit treatment		-0.099	-0.007	-0.045	-0.049	-0.049
		(0.160)	(0.148)	(0.144)	(0.149)	(0.147)
Female leader × Merit treatment		0.191	0.076	0.175	0.182	0.189
		(0.223)	(0.216)	(0.214)	(0.216)	(0.213)
Age	0.057***	0.060***	0.049**	0.044***	0.044**	0.043**
	(0.020)	(0.021)	(0.021)	(0.016)	(0.017)	(0.017)
Number of female siblings	-0.179**	-0.187***	-0.178**	-0.196***	-0.196***	-0.189***
	(0.071)	(0.069)	(0.067)	(0.064)	(0.064)	(0.063)
Number of male siblings	0.017	0.009	-0.077	-0.071	-0.070	-0.065
	(0.082)	(0.082)	(0.069)	(0.068)	(0.069)	(0.069)
Economic status	0.068	0.069	-0.000	-0.020	-0.018	-0.028
	(0.046)	(0.047)	(0.043)	(0.037)	(0.040)	(0.043)
Mother's schooling above father's	0.064	0.054	0.091	0.126	0.127	0.129
	(0.112)	(0.116)	(0.109)	(0.093)	(0.096)	(0.093)
Economics or finance student	0.189	0.189	0.103	-0.001	-0.001	0.027
	(0.119)	(0.118)	(0.111)	(0.109)	(0.110)	(0.116)
Agreeableness index	-0.163	-0.151	-0.201	-0.077	-0.078	-0.074
	(0.278)	(0.286)	(0.289)	(0.290)	(0.291)	(0.284)
All men groups	-0.777***	-0.789***	-0.374	-0.554**	-0.548**	-0.549**
	(0.283)	(0.284)	(0.280)	(0.269)	(0.273)	(0.269)
Correct encryptions during task	0.043	0.042	0.030	0.024	0.023	0.025
	(0.033)	(0.037)	(0.029)	(0.027)	(0.027)	(0.028)
Belief group is better that other group in super-group	0.026	0.061	0.003	0.048	0.047	0.094
	(0.193)	(0.200)	(0.170)	(0.156)	(0.159)	(0.153)
# of times leader believes other players will approve her decision of Alt 2			0.236***	0.198***	0.195***	0.195***
			(0.061)	(0.058)	(0.062)	(0.062)
Preference for equal payment conditions				-0.320**	-0.314**	-0.297**
				(0.121)	(0.131)	(0.135)
Decision to participate in tournament in stage 5					0.022	-0.009
					(0.120)	(0.122)
Willingness to take risks						0.051
						(0.040)
Constant	-1.061*	-1.078	-0.617	-0.248	-0.249	-0.439
	(0.615)	(0.652)	(0.599)	(0.496)	(0.499)	(0.518)
Observations	68	68	68	68	68	68

Notes: OLS regressions. Dependent variable is the preference for tournament payment of leaders. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.13 shows the OLS regressions for leader's selection into a tournament compensation scheme considering the heterogeneity of the sex of the leader and its interaction with group composition. Female leaders in male-majority groups generally select the communal over the agentic compensation, and this is not explained by controls. Oppositely, the choice of the egalitarian payment scheme from female leaders in female-majority groups is explained by altruism and individual preferences on risk and tournament selection.

Table 3.13: OLS regressions of decision of leader to select a tournament compensation (gender composition of groups)

	Stage 6					
	(1)	(2)	(3)	(4)	(5)	(6)
Female leader	-0.340**	-0.420**	-0.360**	-0.455***	-0.454***	-0.420***
	(0.128)	(0.194)	(0.144)	(0.131)	(0.133)	(0.128)
Female majority		0.069	0.055	-0.067	-0.068	-0.057
		(0.152)	(0.145)	(0.135)	(0.135)	(0.138)
Female leader × Female majority		0.057	-0.031	0.247	0.249	0.234
		(0.258)	(0.206)	(0.202)	(0.200)	(0.198)
Merit treatment	0.000	-0.030	0.023	0.017	0.017	0.020
	(0.121)	(0.134)	(0.126)	(0.111)	(0.112)	(0.112)
Age	0.057***	0.055**	0.046**	0.042**	0.042**	0.041**
	(0.021)	(0.022)	(0.021)	(0.017)	(0.017)	(0.017)
Number of female siblings	-0.179**	-0.179**	-0.173**	-0.197***	-0.197***	-0.189***
	(0.071)	(0.072)	(0.068)	(0.063)	(0.064)	(0.063)
Number of male siblings	0.017	0.016	-0.077	-0.054	-0.053	-0.049
	(0.082)	(0.086)	(0.073)	(0.072)	(0.073)	(0.073)
Economic status	0.068	0.063	-0.003	-0.031	-0.030	-0.039
	(0.046)	(0.047)	(0.042)	(0.039)	(0.042)	(0.044)
Mother's schooling above father's	0.064	0.060	0.095	0.128	0.129	0.130
	(0.113)	(0.117)	(0.106)	(0.091)	(0.095)	(0.092)
Economics or finance student	0.189	0.194	0.105	-0.007	-0.007	0.020
	(0.119)	(0.123)	(0.117)	(0.115)	(0.116)	(0.123)
Agreeableness index	-0.163	-0.179	-0.209	-0.102	-0.103	-0.101
	(0.281)	(0.292)	(0.285)	(0.299)	(0.300)	(0.292)
All men groups	-0.777***	-0.777**	-0.344	-0.643**	-0.641**	-0.633**
	(0.286)	(0.330)	(0.300)	(0.296)	(0.303)	(0.301)
Correct encryptions during task	0.043	0.039	0.029	0.025	0.025	0.026
	(0.037)	(0.037)	(0.029)	(0.026)	(0.026)	(0.027)
Belief group is better that other group in super-group	0.026	0.031	-0.006	0.011	0.010	0.053
	(0.197)	(0.198)	(0.167)	(0.152)	(0.153)	(0.150)
# of times leader believes other players will approve her decision of Alt 2			0.238***	0.194***	0.193***	0.194***
			(0.061)	(0.058)	(0.059)	(0.060)
Preference for equal payment conditions				-0.351***	-0.348**	-0.328**
				(0.123)	(0.139)	(0.145)
Decision to participate in tournament in stage 5					0.010	-0.021
					(0.120)	(0.122)
Willingness to take risks						0.048
						(0.040)
Constant	-1.061*	-0.985	-0.572	-0.161	-0.161	-0.340
	(0.626)	(0.669)	(0.590)	(0.492)	(0.495)	(0.512)
Observations	68	68	68	68	68	68
Tests of joint coefficients:						
Additional effect of female leader in FFM		-0.363	-0.390	-0.207	-0.205	-0.186
p-val		$p < 0.05$	$p < 0.05$	0.226	0.214	0.252

Notes: OLS regressions. Dependent variable is the preference for tournament payment of leaders. Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Result 14: Female leaders in male-majority groups are more likely than male leaders in the same group composition to select the egalitarian alternative over the agentic distribution norm. Female leaders and male leaders in female-majority groups are not statistically different in their choice of compensation scheme once including controls.

As women, more than men, are more likely to select an egalitarian compensation, having female leaders appears promising for changes from an agentic to a more communal norm. However, the decision of non-leaders to accept or not this choice determines how much the leader's preference can permeate the group's norm. Table 3.14 shows the OLS estimations for non-leaders where the dependent variables are a dummy that takes the value of 1 for all the cases where non-leaders accept the leader's selection of Alternative 1, regardless of their decision on Alternative 2 (columns (1) to (6)), and a dummy that takes the value of 1 for the cases in which the only acceptable alternative is Alternative 1 (columns (7) to (12)); here, I exclude the non-leaders that accept both potential decisions of the leader. Columns (3), (4), (9) and (10) only include subjects in the random treatment; columns (5), (6), (11) and (12) only include subjects in the merit treatment. To avoid a triple interaction, I present the results of the two treatments separately.

When subjects choose Alternative 2 in Stage 5 they are more likely to accept Alternative 1 in Stage 6, regardless of choice on Alternative 2, if women are leaders in the merit treatment compared to men. However, the additional effect of female leaders in the merit treatment in terms of only accepting Alternative 1, when the DM selected Alternative 2 in Stage 5 is negative, although only marginally significant ($p = 0.164$).

Result 15: Decision-makers who prefer the tournament compensation are less likely to only accept leader's election of an egalitarian norm when the leader is a woman than when the leader is a man.

Table 3.14: OLS regressions of decision of non-leaders to accept egalitarian distribution

	Stage 6							
	Accepts Alt 1 (regardless of Alt 2)				Only accepts leader's decision of Alt 1			
	Random treatment		Merit treatment		Random treatment		Merit treatment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female leader	-0.047	-0.160	-0.013	-0.089	0.101	0.127	-0.076	-0.019
	(0.147)	(0.147)	(0.065)	(0.083)	(0.188)	(0.178)	(0.135)	(0.128)
Decision to participate in tournament in stage 5	0.369*	0.333**	-0.014	-0.018	-0.880***	-0.827***	-0.095	-0.073
	(0.193)	(0.152)	(0.138)	(0.130)	(0.099)	(0.130)	(0.322)	(0.320)
Female leader × Tournament Stage 5	0.157	0.150	0.354**	0.372**	0.063	0.041	-0.514	-0.514
	(0.220)	(0.198)	(0.159)	(0.159)	(0.244)	(0.224)	(0.366)	(0.362)
Female majority	0.093	0.224*	-0.075	-0.019	-0.150	-0.186	0.101	0.055
	(0.107)	(0.119)	(0.100)	(0.093)	(0.155)	(0.144)	(0.205)	(0.193)
Age	0.013	0.021	0.012	0.013	-0.026	-0.032**	0.011	0.010
	(0.023)	(0.023)	(0.021)	(0.019)	(0.016)	(0.015)	(0.027)	(0.024)
Number of female siblings	0.021	0.045	-0.016	-0.006	-0.009	-0.013	-0.025	-0.028
	(0.070)	(0.062)	(0.059)	(0.057)	(0.091)	(0.093)	(0.094)	(0.093)
Number of male siblings	-0.114*	-0.120**	0.093	0.105	0.133***	0.143**	-0.116	-0.141
	(0.062)	(0.049)	(0.066)	(0.067)	(0.048)	(0.053)	(0.080)	(0.086)
Economic status	0.020	0.012	-0.042	-0.030	-0.022	-0.007	0.075	0.055
	(0.054)	(0.058)	(0.036)	(0.030)	(0.064)	(0.080)	(0.064)	(0.054)
Mother's schooling above father's	0.001	0.024	0.051	0.008	-0.082	-0.098	-0.063	0.013
	(0.119)	(0.111)	(0.086)	(0.077)	(0.160)	(0.148)	(0.170)	(0.166)
Economics or finance student	0.074	0.011	0.053	0.023	-0.044	0.003	0.044	0.054
	(0.128)	(0.122)	(0.089)	(0.092)	(0.118)	(0.102)	(0.119)	(0.130)
All men groups	-0.494*	-0.618**						
	(0.277)	(0.265)						
Belief group is better that other group in super-group	-0.116	-0.189*	0.079	0.114	-0.018	0.019	-0.138	-0.175
	(0.117)	(0.107)	(0.067)	(0.082)	(0.089)	(0.104)	(0.122)	(0.133)
Correct encryptions during task	0.001	0.017	0.027	0.037	-0.012	-0.019	-0.034	-0.029
	(0.028)	(0.026)	(0.029)	(0.028)	(0.028)	(0.034)	(0.065)	(0.057)
Preference for equal payment conditions	0.091	0.092	-0.043	-0.049	0.018	0.015	0.013	-0.013
	(0.085)	(0.083)	(0.078)	(0.085)	(0.073)	(0.076)	(0.103)	(0.097)
Willingness to take risks	0.044	0.008	0.008	-0.012	-0.030	-0.019	0.017	0.038
	(0.046)	(0.038)	(0.040)	(0.041)	(0.038)	(0.035)	(0.044)	(0.053)
Female		-0.314***		-0.190		0.120		0.185
		(0.110)		(0.117)		(0.149)		(0.182)
Constant	0.597	0.642	0.636	0.644	1.788***	1.769***	0.829	0.774
	(0.549)	(0.545)	(0.552)	(0.492)	(0.342)	(0.348)	(0.662)	(0.625)
Observations	67	67	68	68	40	40	40	40
Clusters	34	34	34	34	29	29	27	27
Tests of joint coefficients:								
Effect of male leader when competition in Stage 5	0.966	0.975	0.622	0.626	0.908	0.942	0.735	0.701
p-val	$p < 0.1$	$p < 0.1$	0.312	0.256	$p < 0.01$	$p < 0.01$	0.347	0.359
Effect of female leader when competition in Stage 5	1.076	0.966	0.963	0.910	1.072	1.110	0.145	0.168
p-val	$p < 0.1$	0.104	$p < 0.1$	$p < 0.1$	$p < 0.01$	$p < 0.01$	0.811	0.768
Additional effect of female leader over male leader	0.110	-0.010	0.342	0.284	0.164	0.168	-0.590	-0.533
p-val	0.534	0.958	$p < 0.1$	$p < 0.1$	0.385	0.373	0.164	0.197

Notes: OLS regressions. Dependent variables are a dummy that takes the value of 1 for all the cases where non-leaders accept the leader's selection of Alternative 1, regardless of their decision on Alternative 2 (columns (1) to (4)), and a dummy that takes the value of 1 for the cases in which the only acceptable alternative is Alternative 1 (columns (5) to (8)). Columns (1), (2), (5) and (6) only include subjects in the random treatment; columns (3), (4), (7) and (8) only include subjects in the merit treatment. Standard errors clustered at the group level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.15 presents the results considering non-leader's beliefs of the leaders decisions in Stage 6. Tests of joint coefficients are available at the bottom of the table. Believing that male leaders prefer Alternative 1 over Alternative 2 positively and significantly affects the selection of the egalitarian compensation scheme. The effect of believing that a female leader prefers Alternative 1 is also positive and significant, except in the merit treatment. Non-leaders who believe that top performing leaders prefer the egalitarian alternative are more likely to only accept that alternative if the leader is a men than if she is a women. The information on relative performance of at least the women in the group may be causing a belief update on the relative position in the supergroup, that is leading to a preference for a tournament compensation of the non-leaders. The decision to participate in a tournament in Stage 5 negatively affects the acceptance of Alternative 1 only, except in the merit treatment.

Table 3.15: OLS regressions of decision of non-leaders to only accept egalitarian distribution, including beliefs of leader's preferences

	Stage 6					
	Random treatment			Merit treatment		
	(1)	(2)	(3)	(4)	(5)	(6)
Belief leader's preference for Alt 1	0.210	0.235	0.285	0.003	-0.001	-0.007
	(0.234)	(0.227)	(0.247)	(0.224)	(0.221)	(0.196)
Female leader	0.011	0.129	0.231	-0.530***	-0.545***	-0.493**
	(0.181)	(0.232)	(0.263)	(0.170)	(0.190)	(0.192)
Belief Alt 1 × Female leader	0.020	-0.065	-0.161	0.612***	0.618***	0.624***
	(0.218)	(0.246)	(0.282)	(0.180)	(0.183)	(0.178)
Age	-0.022*	-0.032*	-0.045**	0.006	0.006	0.004
	(0.012)	(0.018)	(0.020)	(0.026)	(0.027)	(0.024)
Number of female siblings	-0.009	-0.017	-0.021	0.008	0.008	0.004
	(0.097)	(0.095)	(0.091)	(0.082)	(0.085)	(0.085)
Number of male siblings	0.103*	0.095	0.114**	-0.079	-0.080	-0.105
	(0.055)	(0.058)	(0.055)	(0.064)	(0.066)	(0.073)
Economic status	-0.012	-0.032	0.006	0.091	0.090*	0.070
	(0.033)	(0.043)	(0.064)	(0.054)	(0.052)	(0.045)
Mother's schooling above father's	-0.074	-0.076	-0.094	-0.142	-0.139	-0.063
	(0.119)	(0.124)	(0.115)	(0.142)	(0.132)	(0.118)
Economics or finance student	0.012	-0.042	0.046	0.049	0.052	0.061
	(0.130)	(0.110)	(0.112)	(0.085)	(0.100)	(0.110)
Belief group is better that other group in super-group	-0.055	-0.034	0.035	-0.326*	-0.325*	-0.363**
	(0.113)	(0.096)	(0.098)	(0.184)	(0.189)	(0.166)
Correct encryptions during task	-0.038	-0.025	-0.037	-0.088	-0.088	-0.083
	(0.029)	(0.026)	(0.031)	(0.061)	(0.062)	(0.050)
Preference for equal payment conditions	0.010	0.002	-0.018	0.051	0.049	0.023
	(0.095)	(0.095)	(0.100)	(0.093)	(0.093)	(0.087)
Decision to participate in tournament in stage 5	-0.765***	-0.814***	-0.745***	-0.216	-0.212	-0.193
	(0.195)	(0.198)	(0.228)	(0.215)	(0.224)	(0.236)
Willingness to take risks	-0.039	-0.033	-0.011	0.037	0.037	0.059
	(0.042)	(0.046)	(0.046)	(0.038)	(0.040)	(0.047)
Female majority		-0.151	-0.226		0.019	-0.026
		(0.145)	(0.160)		(0.150)	(0.136)
Female			0.229			0.186
			(0.191)			(0.170)
Constant	1.709***	1.929***	1.894***	1.447*	1.438*	1.384*
	(0.281)	(0.401)	(0.402)	(0.772)	(0.809)	(0.713)
Observations	40	40	40	40	40	40
Clusters	29	29	29	27	27	27
Tests of joint coefficients:						
Additional effect of male leader choosing Alt 1	1.719	1.929	1.894	1.447	1.438	1.384
p-val	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.1$	$p < 0.1$	$p < 0.1$
Additional effect of female leader choosing Alt 1	1.720	2.058	2.125	0.918	0.893	0.891
p-val	$p < 0.01$	$p < 0.01$	$p < 0.01$	0.290	0.337	0.255
Additional effect of female leader	0.011	0.129	0.231	-0.530	-0.545	-0.493
p-val	0.950	0.581	0.387	$p < 0.01$	$p < 0.01$	$p < 0.05$

Notes: OLS regressions. Dependent variable considers subjects that decided for only one alternative, and takes the value of 1 for the cases in which that alternative was Alternative 1. Columns (1) to (3) only include subjects in the random treatment; columns (4) to (6) only include subjects in the merit treatment. Standard errors clustered at the group level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Last, Table 3.16 keeps only subjects who chose Alternative 2 in Stage 5 and accept the leader's decision to select Alternative 1 in Stage 6. It considers those willing to change their preferred norm after the leader. Women are more likely than men to accept an alternative against their preference when a leader decides it, particularly in merit treatment. However, the additional effect of women accepting the choice of Alternative 1 of a female leader over a male leader is negative and significant, particularly in the merit treatment. Men in the merit treatment also accept more frequently the selection of the communal norm only when a male leader decides it. Consequently, female leaders are less influential than male leaders in attaining an egalitarian rule, mainly if the leader is deserving of that role.

Result 16: Female leaders are less influential than male leaders to promote changes towards an egalitarian rule, mainly in the merit treatment.

Table 3.16: OLS regressions of decision of non-leaders who prefer Alternative 2 to accept egalitarian distribution

	Stage 6					
	Random treatment			Merit treatment		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.559*	0.519	0.520*	1.000**	1.001*	0.876
	(0.281)	(0.304)	(0.256)	(0.470)	(0.519)	(0.528)
Female leader	-0.071	-0.084	0.081	0.465	0.466	0.036
	(0.264)	(0.269)	(0.244)	(0.563)	(0.597)	(0.654)
Female × Female leader	-0.301	-0.328	0.067	-1.034	-1.034	-0.918
	(0.323)	(0.325)	(0.389)	(0.759)	(0.810)	(0.806)
Belief group is better that other group in super-group	0.617	0.636	0.832**	-0.942	-0.944	-1.054
	(0.406)	(0.403)	(0.356)	(0.556)	(0.686)	(0.726)
Correct encryptions during task	-0.037	-0.028	-0.046	-0.230*	-0.230*	-0.228
	(0.098)	(0.102)	(0.100)	(0.118)	(0.127)	(0.138)
Age	0.017	0.023	-0.011	-0.127*	-0.127*	-0.126*
	(0.052)	(0.055)	(0.039)	(0.066)	(0.073)	(0.073)
Number of female siblings	-0.188**	-0.185**	-0.212***	0.116	0.116	0.109
	(0.071)	(0.073)	(0.060)	(0.124)	(0.137)	(0.137)
Number of male siblings	0.125	0.123	0.099	-0.330**	-0.331**	-0.391**
	(0.080)	(0.086)	(0.068)	(0.143)	(0.151)	(0.144)
Economic status	-0.033	-0.021	-0.061	0.027	0.027	0.039
	(0.099)	(0.105)	(0.089)	(0.089)	(0.092)	(0.099)
Mother's schooling above father's	0.140	0.163	-0.029	0.228	0.228	0.312
	(0.184)	(0.190)	(0.171)	(0.208)	(0.219)	(0.264)
Economics or finance student	0.098	0.070	0.186	-0.437	-0.437	-0.493
	(0.407)	(0.416)	(0.366)	(0.315)	(0.336)	(0.359)
All men groups	0.995**	0.936*	0.971**			
	(0.474)	(0.492)	(0.418)			
Preference for equal payment conditions	-0.198	-0.276	-0.129	0.299	0.300	0.364
	(0.353)	(0.381)	(0.290)	(0.313)	(0.343)	(0.355)
Decision to participate in tournament in stage 3	-0.548**	-0.546**	-0.603***	0.253	0.253	0.227
	(0.211)	(0.206)	(0.210)	(0.225)	(0.231)	(0.233)
Willingness to take risks		-0.049	0.032		0.001	0.056
		(0.052)	(0.065)		(0.090)	(0.106)
Female majority		-0.547**			0.473	
			(0.255)			(0.382)
Constant	0.193	0.112	0.753	5.492**	5.495**	5.335**
	(1.079)	(1.168)	(0.826)	(2.188)	(2.389)	(2.480)
Observations	30	30	30	29	29	29
Clusters	23	23	23	23	23	23
Tests of joint coefficients:						
Additional effect of female leader for female vs male DM	0.257	0.191	0.587	-0.033	-0.033	-0.042
p-val	0.377	0.492	$p < 0.1$	0.932	0.934	0.914
Additional effect of female vs male leader if female DM	-0.373	-0.412	0.149	-0.568	-0.568	-0.882
p-val	0.173	0.147	0.725	$p < 0.1$	$p < 0.1$	$p < 0.05$

Notes: OLS regressions. Dependent variable considers subjects that select Alternative 2 in Stage 5 and accept Alternative 1 in Stage 6. Columns (1) to (3) only include subjects in the random treatment; columns (41) to (6) only include subjects in the merit treatment. Standard errors clustered at the group level in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

3.4 Discussion and conclusions

I propose an experimental design to study the role of female or male leaders for the acceptance of an egalitarian norm over one in which some obtain gains by sacrificing others. I argue that increasing the valuation of communal arrangements can help close the gender gaps in time use on non-promotable tasks, which are critical to the organizational objectives and disproportionally rely on women. I separate from the literature that promotes changes in women's attitudes toward competition or in the institutions in which competition happens as the way to close the gender gaps in labor market outcomes. In my experimental design, I introduce a female or male leader who decides the payment norm for the group. I also include an exogenous variation on why the leader is selected in that role (meritocratic or non-meritocratic) and vary the gender composition of the group.

After confirming results at the individual level, I find that female-majority groups unconditionally favor the selection of an egalitarian rule, particularly for women who are part of that group. Furthermore, women in these groups are less likely to change from an individual egalitarian preference to a group's agentic norm, which is robust to beliefs, performance, controls, altruism, individual preferences for competition, and risk. Oppositely, no particular group configuration favors men selecting a communal payment scheme, and only the most altruistic men in either group composition select an egalitarian payment over an agentic one. An unexplored issue is how these results change with single-sex groups and when the interactions happen between groups of different gender compositions.

My main contribution is, however, in terms of the acceptance of the communal norm with female and male leaders. I find that female leaders select tournament compensation less often than male leaders. Women's decisions as leaders appear to be affected by how they expect other group members to react to their compensation choices. Furthermore, female leaders in male-majority groups appear to select the communal norm unconditionally. In contrast, the same decision for women in female-majority groups depends on altruism, risk, and individual preferences for competition. Non-leaders who prefer the tournament compensation, both at the individual level and as part of a group, are more likely to support the leader's decision of the egalitarian norm when the leader is male rather than female. In addition, the belief that the leader prefers the egalitarian norm influences the support of that norm only for male leaders. That is, non-leaders care about supporting the leader's preferences only in the case when those leaders are men. Last, female leaders have less influence on changing individual behavior toward a communal norm than male leaders, particularly in the merit treatment. This paper does not explore how having a stereotypically male task could change these results. This is an open question for future research.

These results appear consistent with the finding that women in a position of power are judged more negatively than men (Chakraborty and Serra, 2022; Abel, 2022; Boring, 2017; Grossman et al., 2019). The fact that male leaders have the most legitimacy to change the social norm toward commonality, and are the ones who value this norm the least, represents a dilemma. The difficulty of implementing the required institutional changes increases as male leaders are more likely to occupy leadership positions, and that the expectation from leaders is that they exhibit agentic attitudes. The discussion on how to promote the valuation of communal norms is open to further research.

Bibliography

Abel, M. (2022). Do workers discriminate against female bosses? *Journal of Human Resources*, pages 1120–11318R3.

Adams, R. B. and Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management science*, 58(2):219–235.

Ahlert, M., Felder, S., and Vogt, B. (2012). Which patients do i treat? an experimental study with economists and physicians. *Health Economics Review*, 2(1):1–11.

Ahlert, M. and Funke, K. (2012). *A mental model for decision making in allocating a medical resource*. Jacobs Univ., FOR 655.

Ahlert, M., Funke, K., and Schwettmann, L. (2013). Thresholds, productivity, and context: an experimental study on determinants of distributive behaviour. *Social Choice and Welfare*, 40(4):957–984.

Ahlert, M. and Schwettmann, L. (2017). Allocating health care resources: a questionnaire experiment on the predictive success of rules. *International Journal for Equity in Health*, 16(1):112.

Apesteguia, J., Azmat, G., and Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–47.

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

Balafoutas, L., Kerschbamer, R., and Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of economic behavior & organization*, 83(1):125–135.

Bartling, B., Fehr, E., Maréchal, M. A., and Schunk, D. (2009). Egalitarianism and competitiveness. *American Economic Review*, 99(2):93–98.

Beeney, J. E., Franklin Jr, R. G., Levy, K. N., and Adams Jr, R. B. (2011). I feel your pain: emotional closeness modulates neural responses to empathically experienced rejection. *Social Neuroscience*, 6(4):369–376.

Bejarano, H., Green, E. P., and Rassenti, S. (2017). Payment scheme self-selection in the credence goods market: An experimental study. *Journal of Economic Behavior & Organization*, 142:396–403.

Benndorf, V., Rau, H. A., and Sölch, C. (2019). Minimizing learning in repeated real-effort tasks. *Journal of Behavioral and Experimental Finance*, 22:239–248.

Berniell, L., de la Mata, D., Bernal, R., Camacho, A., Barrera-Osorio, F., Álvarez, F., Brassiolo, P., and Vargas, J. (2016). Red 2016. más habilidades para el trabajo y la vida: los aportes de la familia, la escuela, el entorno y el mundo laboral.

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

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of public economics*, 145:27–41.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Brañas-Garza, P., Capraro, V., and Rascon-Ramirez, E. (2018). Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters*, 170:19–23.

Brendel, F., Einhaus, L., and Then, F. (2021). Resource scarcity and prioritization decisions in medical care: A lab experiment with heterogeneous patient types. *Health Economics*, 30(2):470–477.

Brock, J. M., Lange, A., and Leonard, K. L. (2016). Generosity and prosocial behavior in healthcare provision evidence from the laboratory and field. *Journal of Human Resources*, 51(1):133–162.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., and Wiesen, D. (2013). *How Effective are Pay-for-Performance Incentives for Physicians?–A Laboratory Experiment.* Number 413. Ruhr Economic Papers.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., Wiesen, D., et al. (2019). Physician performance pay: Experimental evidence. *SSRN Electron. J*, 3467583.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2016). Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior & Organization*, 131:17–23.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, 26(2):243–262.

BSR (2021). Here to be heard.

Buckles, K. (2019). Fixing the leaky pipeline: Strategies for making economics work for women at every stage. *Journal of economic perspectives*, 33(1):43–60.

Byron, K. and Post, C. (2016). Women on boards of directors and corporate social performance: A meta-analysis. *Corporate Governance: An International Review*, 24(4):428–442.

Cassar, A. and Rigdon, M. L. (2021a). Option to cooperate increases women's competitiveness and closes the gender gap. *Evolution and Human Behavior*, 42(6):556–572.

Cassar, A. and Rigdon, M. L. (2021b). Prosocial option increases women's entry into competition. *Proceedings of the National Academy of Sciences*, 118(45):e2111943118.

Cassar, A. and Zhang, Y. J. (2022). The competitive woman evolutionary insights and cross-cultural evidence into finding the femina economica. *Journal of Economic Behavior & Organization*, 197:447–471.

Chakraborty, P. and Serra, D. (2022). Gender and leadership in organizations: The threat of backlash. Technical report, Working Paper.

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

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Choné, P. and Ma, C.-T. A. (2011). Optimal health care contract under physician agency. *Annals of Economics and Statistics*, 101(102):229–256.

Coffman, K. B. and Roth, A. E. (2021). Do women shy away from competition? do men compete too much?(by muriel niederle and lise vesterlund). In *The Art of Experimental Economics*, pages 198–209. Routledge.

Cristofori, I., Harquel, S., Isnard, J., Mauguière, F., and Sirigu, A. (2015). Monetary reward suppresses anterior insula activity during social pain. *Social cognitive and affective neuroscience*, 10(12):1668–1676.

Dasgupta, U., Mani, S., Sharma, S., and Singhal, S. (2019). Can gender differences in distributional preferences explain gender gaps in competition? *Journal of Economic Psychology*, 70:1–11.

Di Guida, S., Gyrd-Hansen, D., and Oxholm, A. S. (2019). Testing the myth of fee-for-service and overprovision in health care. *Health Economics*, 28(5):717–722.

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

Doñate-Buendía, A., García-Gallego, A., and Petrović, M. (2022). Gender and other moderators of giving in the dictator game: A meta-analysis. *Journal of Economic Behavior & Organization*, 198:280–301.

Drouvelis, M. and Rigdon, M. L. (2022). Gender differences in competitiveness: The role of social incentives. *Available at SSRN 4009418*.

Duch, M., Grossmann, M. R., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Working paper*.

Dufwenberg, M. and Muren, A. (2006). Gender composition in teams. *Journal of Economic Behavior & Organization*, 61(1):50–54.

Dusheiko, M., Gravelle, H., Jacobs, R., and Smith, P. (2006). The effect of financial incentives on gatekeeping doctors: evidence from a natural experiment. *Journal of Health Economics*, 25(3):449–478.

Eisenberger, N. I., Lieberman, M. D., and Williams, K. D. (2003). Does rejection hurt? an fmri study of social exclusion. *Science*, 302(5643):290–292.

Ellis, R. P. and McGuire, T. G. (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics*, 5(2):129–151.

Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fornwagner, H., Pompeo, M., and Serdarevic, N. (2020). Him or her? choosing competition on behalf of someone else. *Choosing competition on behalf of someone else (June 4, 2020)*.

Godager, G., Hennig-Schmidt, H., and Iversen, T. (2016). Does performance disclosure influence physicians' medical decisions? an experimental study. *Journal of Economic Behavior & Organization*, 131:36–46.

Gonsalkorale, K. and Williams, K. D. (2007). The kkk won't let me play: Ostracism even by a despised outgroup hurts. *European Journal of Social Psychology*, 37(6):1176–1186.

Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization*, 106:367–378.

Greiner, B. et al. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working paper series in economics*, 10(23):63–104.

Grossman, P. J., Eckel, C., Komai, M., and Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior & Organization*, 161:197–215.

Grossman, P. J., Komai, M., and Jensen, J. E. (2015). Leadership and gender in groups: An experiment. *Canadian Journal of Economics/Revue canadienne d'économique*, 48(1):368–388.

Gutierrez, I. A., Molina, O., and Ñopo, H. (2018). Stand against bullying: An experimental school intervention.

Hafner, L., Reif, S., and Seebauer, M. (2017). Physician behavior under prospective payment schemes: Evidence from artefactual field and lab experiments. Technical report, FAU Discussion Papers in Economics.

Harter, C. L., Becker, W. E., and Watts, M. (2011). Time allocations and reward structures for us academic economists from 1995–2005: Evidence from three national surveys. *International Review of Economics Education*, 10(2):6–27.

Healy, A. and Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, 121(555):1192–1204.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, 6(1):689–733.

Hennig-Schmidt, H., Selten, R., and Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, 30(4):637–646.

Houser, D. and Schunk, D. (2009). Social environments with competitive pressure: Gender effects in the decisions of german schoolchildren. *Journal of Economic Psychology*, 30(4):634–641.

John, O. P., Donahue, E. M., and Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

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

Keser, C., Montmarquette, C., Schmidt, M., and Schnitzler, C. (2020). Custom-made health-care: an experimental investigation. *Health Economics Review*, 10(1):1–12.

Keser, C., Peterle, E., and Schnitzler, C. (2014). Money talks-paying physicians for performance. *Revised Version, Cege Discussion Paper*, (173).

Kessel, D., Mollerstrom, J., and Van Veldhuizen, R. (2021). Can simple advice eliminate the gender gap in willingness to compete? *European Economic Review*, 138:103777.

Kirsch, A. (2018). The gender composition of corporate boards: A review and research agenda. *The Leadership Quarterly*, 29(2):346–364.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lagarde, M. and Blaauw, D. (2017). Physicians' responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine*, 179:147–159.

Lereya, S. T., Copeland, W. E., Costello, E. J., and Wolke, D. (2015). Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *The Lancet Psychiatry*, 2(6):524–531.

Lu, J. and Herremans, I. M. (2019). Board gender diversity and environmental performance: An industries perspective. *Business Strategy and the Environment*, 28(7):1449–1464.

Manchester, C. and Barbezat, D. (2013). The effect of time use in explaining male–female productivity differences among economists. *Industrial Relations: A Journal of Economy and Society*, 52(1):53–77.

Markowsky, E. and Beblo, M. (2022). When do we observe a gender gap in competition entry? a meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, 198:139–163.

Martinsson, P. and Persson, E. (2019). Physician behavior and conditional altruism: the effects of payment system and uncertain health benefit. *Theory and Decision*, 87(3):365–387.

Masten, C. L., Eisenberger, N. I., Pfeifer, J. H., and Dapretto, M. (2010). Witnessing peer rejection during early adolescence: Neural correlates of empathy for experiences of social exclusion. *Social neuroscience*, 5(5-6):496–507.

Masten, C. L., Morelli, S. A., and Eisenberger, N. I. (2011). An fmri investigation of empathy for 'social pain'and subsequent prosocial behavior. *Neuroimage*, 55(1):381–388.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

Niederle, M. (2020). Gender. In Kagel, J. H. and Roth, A. E., editors, *The handbook of experimental economics, volume 2*, pages 481–209. Princeton university press.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101.

Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European journal of psychology of education*, 12(4):495–510.

Oxholm, A. S., Di Guida, S., and Gyrd-Hansen, D. (2021). Allocation of health care under pay for performance: Winners and losers. *Social Science & Medicine*, 278:1–9.

Oxholm, A. S., Di Guida, S., Gyrd-Hansen, D., and Olsen, K. R. (2019). Taking care of high-need patients in capitation-based payment schemes–an experimental investigation into the importance of market conditions. *Applied Economics*, 51(47):5174–5184.

Polanin, J. R., Espelage, D. L., and Pigott, T. D. (2012). A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *School Psychology Review*, 41(1):47–65.

Riem, M. M., Bakermans-Kranenburg, M. J., Huffmeijer, R., and van IJzendoorn, M. H. (2013). Does intranasal oxytocin promote prosocial behavior to an excluded fellow player? a randomized-controlled trial with cyberball. *Psychoneuroendocrinology*, 38(8):1418–1425.

Rivara, F. and Le Menestrel, S. (2016). *Preventing bullying through science, policy, and practice*. National Academies Press.

Rudman, L. A. and Glick, P. (2021). *The social psychology of gender: How power and intimacy shape gender relations*, chapter 1,4,5. Guilford Publications.

Salgado, E., Vargas-Trujillo, E., Schmutzler, J., and Wills-Herrera, E. (2016). Uso del inventario de los cinco grandes en una muestra colombiana. *Avances en Psicología Latinoamericana*, 34(2):365–382.

Sarzosa, M. (2021). Victimization and skill accumulation: The case of school bullying. *Journal of Human Resources*, pages 0819–10371R2.

Sarzosa, M. and Urzúa, S. (2021). Bullying among adolescents: The role of skills. *Quantitative Economics*, 12(3):945–980.

Sebastian, C., Viding, E., Williams, K. D., and Blakemore, S.-J. (2010). Social brain development and the affective consequences of ostracism in adolescence. *Brain and cognition*, 72(1):134–145.

Shen, W. and Joseph, D. L. (2021). Gender and leadership: A criterion-focused review and research agenda. *Human Resource Management Review*, 31(2):100765.

Smith, P. K. and Brain, P. (2000). Bullying in schools: Lessons from two decades of research. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 26(1):1–9.

Tarazona-Gomez, M. (2004). Are individuals prudent? an experimental approach using lottery choices. *Lerna-Ehess Working Paper*.

Tate, G. and Yang, L. (2015). Female leadership and gender equity: Evidence from plant closure. *Journal of Financial Economics*, 117(1):77–97.

Taylor, S. W., Fender, B. F., and Burke, K. G. (2006). Unraveling the academic productivity of economists: The opportunity costs of teaching and service. *Southern Economic Journal*, 72(4):846–859.

Vrijhof, C. I., van den Bulk, B. G., Overgaauw, S., Lelieveld, G.-J., Engels, R. C., and van IJzendoorn, M. H. (2016). The prosocial cyberball game: compensating for social exclusion and its associations with empathic concern and bullying in adolescents. *Journal of adolescence*, 52:27–36.

Williams, K. D., Cheung, C. K., and Choi, W. (2000). Cyberostracism: effects of being ignored over the internet. *Journal of personality and social psychology*, 79(5):748.

Williams, K. D. and Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior research methods*, 38(1):174–180.

Zhang, P., Zhang, Y., and Palma, M. (2020). Social norms and competitiveness: My willingness to compete depends on who i am (supposed to be). *Available at SSRN 4010949*.

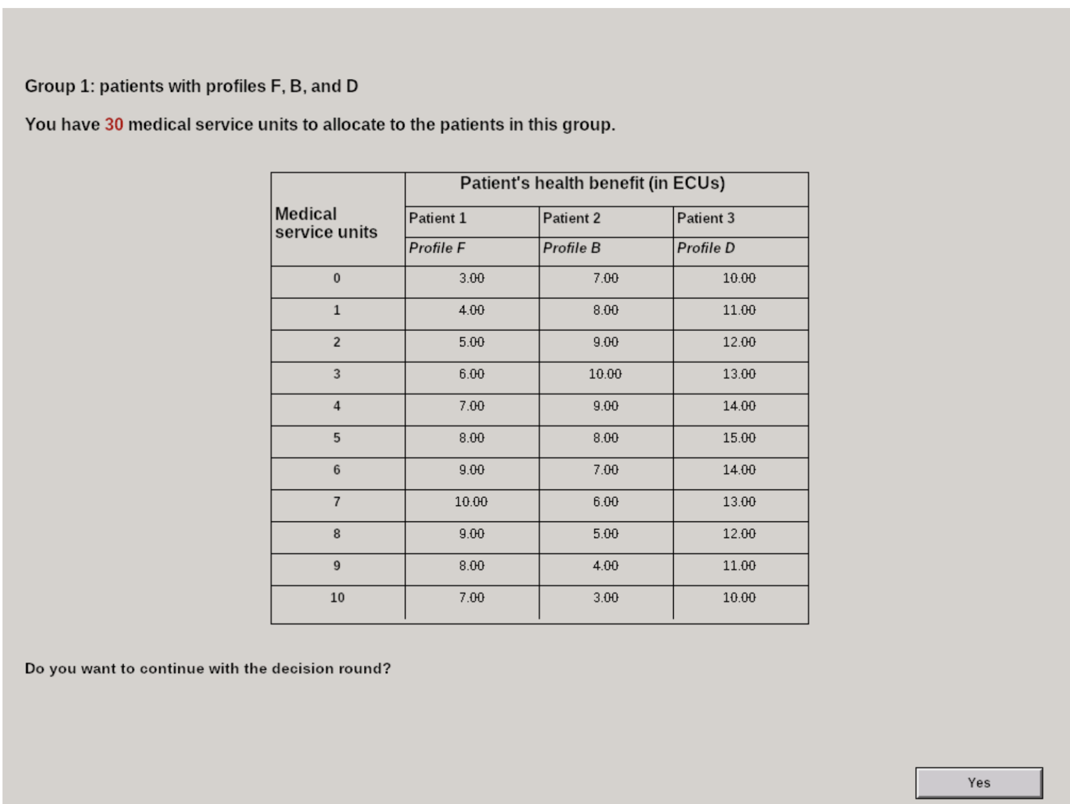
Zhao, S., López Vargas, K., Friedman, D., and Gutierrez, M. (2020). Ucsc leeps lab protocol for online economics experiments. *Available at SSRN 3594027*.

Appendices

Appendix A

Chapter 1

Recall that participants observed labels F instead of A and vice versa and C instead of E and vice versa.



Group 1: patients with profiles F, B, and D

You have 30 medical service units to allocate to the patients in this group.

Medical service units	Patient's health benefit (in ECUs)		
	Patient 1	Patient 2	Patient 3
	Profile F	*Profile B*	*Profile D*
0	3.00	7.00	10.00
1	4.00	8.00	11.00
2	5.00	9.00	12.00
3	6.00	10.00	13.00
4	7.00	9.00	14.00
5	8.00	8.00	15.00
6	9.00	7.00	14.00
7	10.00	6.00	13.00
8	9.00	5.00	12.00
9	8.00	4.00	11.00
10	7.00	3.00	10.00

Do you want to continue with the decision round?

Yes

Figure A.1: Group Screen - Control

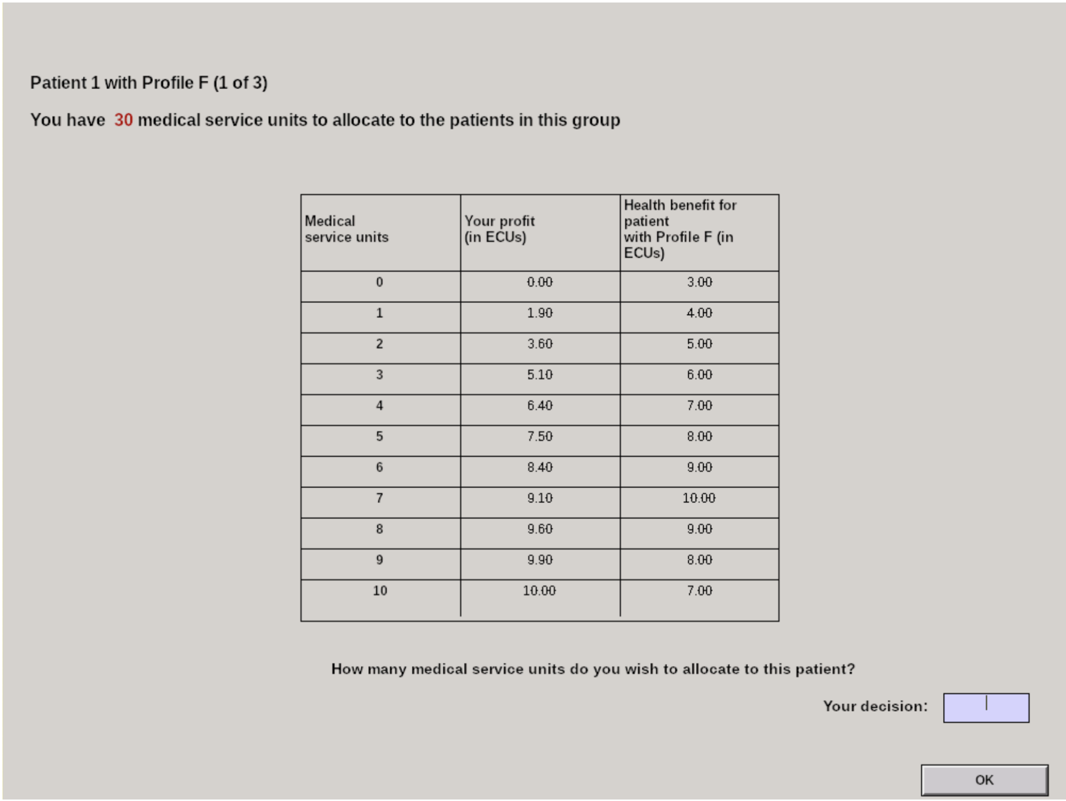
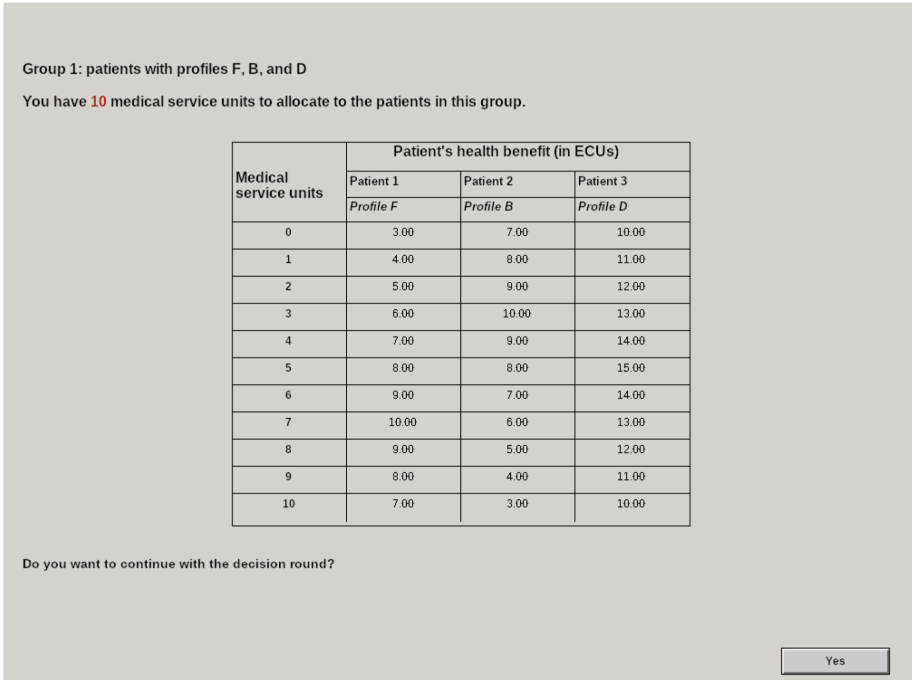


Figure A.2: Patient Screen - Control

Table A.1: Physicians by configuration

BC								**U and BC+U**			
1		**2**		**3**		**4**		**1 or 3**		**2 or 4**	
Part 1	**Part 2**	**Part 1**	**Part 2**	**Part 1**	**Part 2**	**Part 1**	**Part 2**	**Part 1**	**Part 2**	**Part 1**	**Part 2**
ABD	ABD	ABC	ABC	ABD	ABC	ABC	ABD	ABD	AB D/C	ABC	AB D/C
DGC	DGC	DGF	DGF	DGC	DGF	DGF	DGC	DGC	DG C/F	DGF	DG C/F
FEG	FEG	FED	FED	FEG	FED	FED	FEG	FEG	FE G/D	FED	FE G/D

Notes: BC stands for Budget Constraint, U for Uncertainty and BC+U to Budget Constraint and Uncertainty treatments.



Group 1: patients with profiles F, B, and D

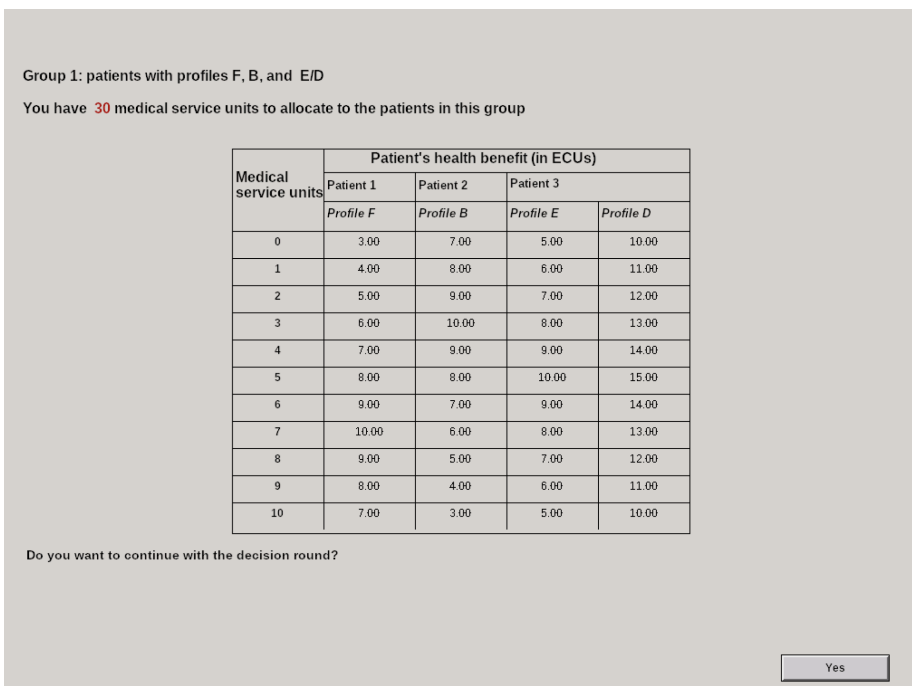
You have 10 medical service units to allocate to the patients in this group.

Medical service units	Patient's health benefit (in ECUs)		
	Patient 1	Patient 2	Patient 3
	Profile F	*Profile B*	*Profile D*
0	3.00	7.00	10.00
1	4.00	8.00	11.00
2	5.00	9.00	12.00
3	6.00	10.00	13.00
4	7.00	9.00	14.00
5	8.00	8.00	15.00
6	9.00	7.00	14.00
7	10.00	6.00	13.00
8	9.00	5.00	12.00
9	8.00	4.00	11.00
10	7.00	3.00	10.00

Do you want to continue with the decision round?

Yes

Figure A.3: Group Screen - BC Treatment



Group 1: patients with profiles F, B, and E/D

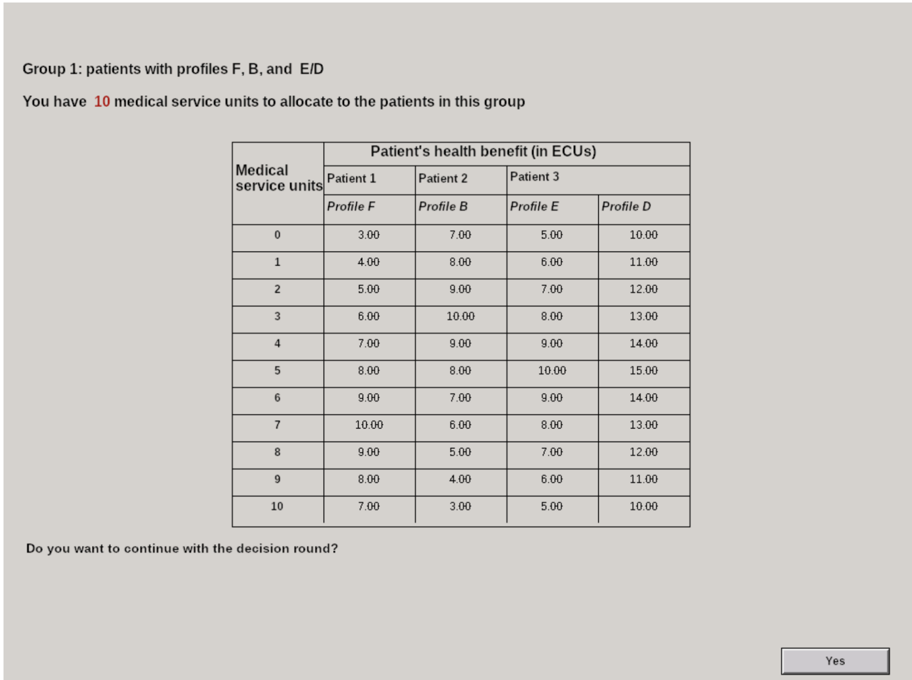
You have 30 medical service units to allocate to the patients in this group

Medical service units	Patient's health benefit (in ECUs)			
	Patient 1	Patient 2	Patient 3	
	Profile F	*Profile B*	*Profile E*	*Profile D*
0	3.00	7.00	5.00	10.00
1	4.00	8.00	6.00	11.00
2	5.00	9.00	7.00	12.00
3	6.00	10.00	8.00	13.00
4	7.00	9.00	9.00	14.00
5	8.00	8.00	10.00	15.00
6	9.00	7.00	9.00	14.00
7	10.00	6.00	8.00	13.00
8	9.00	5.00	7.00	12.00
9	8.00	4.00	6.00	11.00
10	7.00	3.00	5.00	10.00

Do you want to continue with the decision round?

Yes

Figure A.4: Group Screen - U Treatment



Group 1: patients with profiles F, B, and E/D

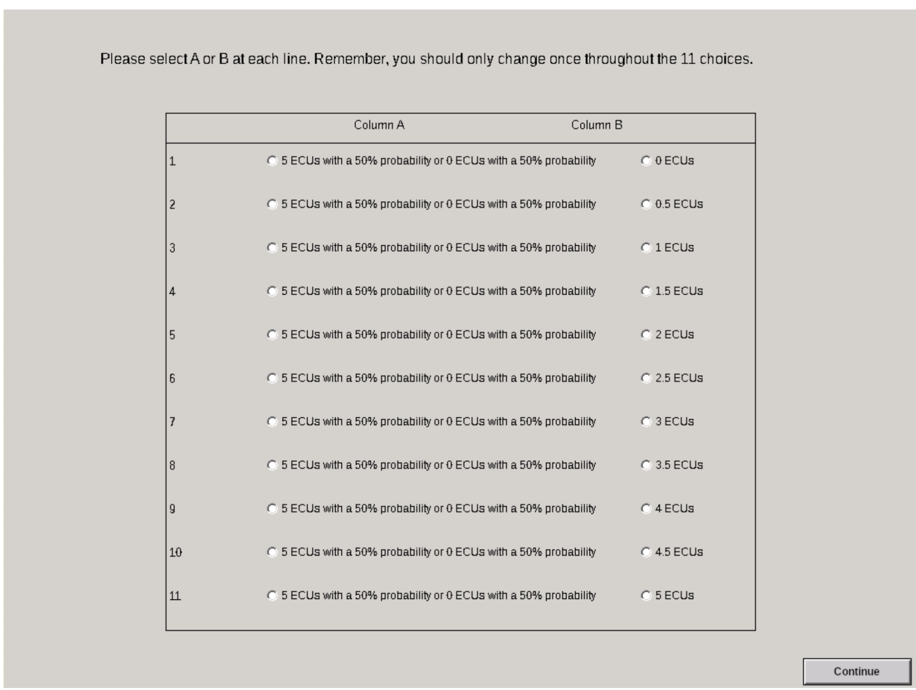
You have 10 medical service units to allocate to the patients in this group

Medical service units	Patient's health benefit (in ECUs)			
	Patient 1	Patient 2	Patient 3	
	Profile F	*Profile B*	*Profile E*	*Profile D*
0	3.00	7.00	5.00	10.00
1	4.00	8.00	6.00	11.00
2	5.00	9.00	7.00	12.00
3	6.00	10.00	8.00	13.00
4	7.00	9.00	9.00	14.00
5	8.00	8.00	10.00	15.00
6	9.00	7.00	9.00	14.00
7	10.00	6.00	8.00	13.00
8	9.00	5.00	7.00	12.00
9	8.00	4.00	6.00	11.00
10	7.00	3.00	5.00	10.00

Do you want to continue with the decision round?

Yes

Figure A.5: Patient Screen - BC+U Treatment



Please select A or B at each line. Remember, you should only change once throughout the 11 choices.

	Column A	Column B
1	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	0 ECUs
2	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	0.5 ECUs
3	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	1 ECUs
4	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	1.5 ECUs
5	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	2 ECUs
6	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	2.5 ECUs
7	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	3 ECUs
8	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	3.5 ECUs
9	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	4 ECUs
10	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	4.5 ECUs
11	5 ECUs with a 50% probability or 0 ECUs with a 50% probability	5 ECUs

Continue

Figure A.6: Lottery to measure risk aversion preference

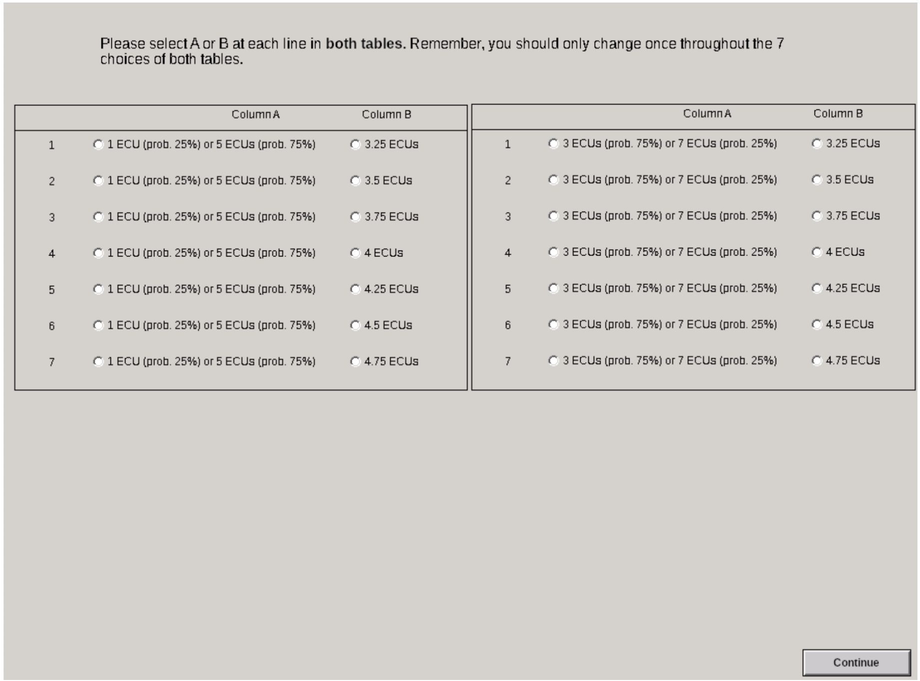


Figure A.7: Lotteries to measure prudence preference

Table A.2: Summary statistics and mean comparison tests, by treatment

	Mean			p-value		
	BC	**U**	**BC+U**	**BC vs U**	**BC vs BC+U**	**U vs BC+U**
Age	21.10	21.88	21.43	0.103	0.488	0.321
Female	0.63	0.65	0.57	0.783	0.549	0.379
Siblings	1.59	2.02	1.31	0.068	0.203	0.002
Economics	0.12	0.12	0.12	0.972	1.000	0.972
Econ or Finance	0.29	0.23	0.27	0.470	0.828	0.613
Medicine	0.02	0.04	0.02	0.572	1.000	0.572
Med or Physiotherapy	0.20	0.25	0.22	0.515	0.809	0.684
Family Income at 18	3.41	3.37	3.18	0.793	0.185	0.256
Med Parents	0.14	0.10	0.16	0.521	0.782	0.359
Socioeconomic status	3.63	3.65	3.14	0.912	0.028	0.030
Self-financed Expense	3.36	3.70	4.49	0.642	0.126	0.295
Risk aversion	0.20	0.24	0.22	0.64	0.86	0.79
Altruism	1.68	2.07	1.96	0.03	0.09	0.54
Prudence	0.31	0.35	0.47	0.73	0.11	0.20

Notes: This table shows the mean by treatment for the main final survey variables and the p-value for the mean differences by treatment pair. BC: Budget Constraint; U: Uncertainty; BC+U: Budget Constraint and Uncertainty. We present results for age, gender, number of siblings, and if the participant is an Economics, Finance, Medicine or Physiotherapy major. We also add socioeconomic status following the Colombian classification. Our measure of risk aversion is a dummy that takes the value of 1 for risk-averse subjects; this measure excludes those with multiple switching points. Similarly, our measure of prudence takes the value of 1 for subjects classified as prudent, and 0 for those classified as imprudent or neutral following Tarazona-Gomez (2004). Finally, our measure of altruism considers the number of EMUs sent from one player to another.

Appendix B

Chapter 2

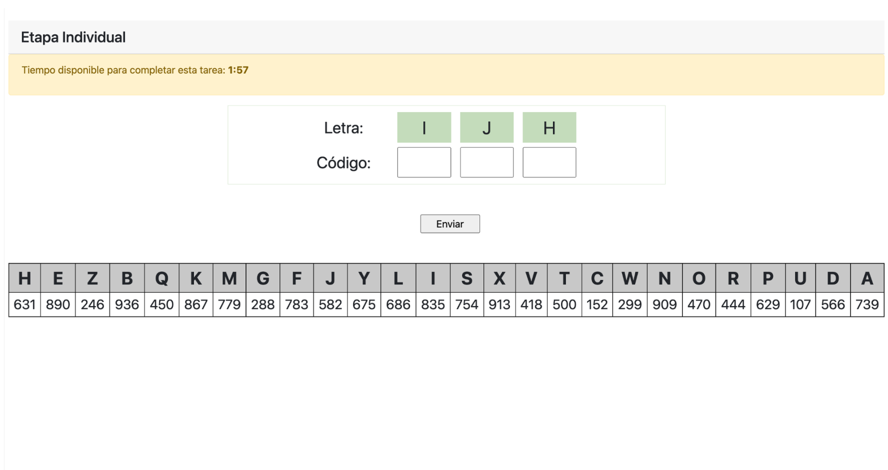


Figure B.1: Individual stage RET screen

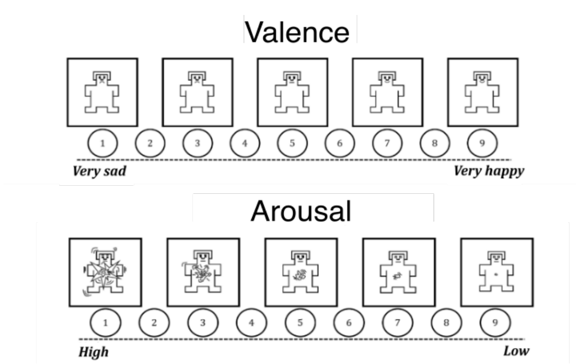


Figure B.2: Images from the Emotion Test

Table B.1: OLS results for within-subject changes in normalized RET performance by exclusion, by role. Detailed table

	$\frac{RET_2 - RET_1}{RET_1}$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Potential Offender					*Potential Victim*				
Less than 3 exclusions by part	-0.025	-0.008	-0.017	-0.016	-0.017	-0.041	-0.058	-0.063*	-0.066*	-0.063*
	(0.064)	(0.075)	(0.061)	(0.062)	(0.063)	(0.038)	(0.035)	(0.035)	(0.039)	(0.035)
3 or more exclusions by part	0.039	0.046	0.037	0.042	0.040	0.037	-0.012	-0.023	-0.022	-0.023
	(0.064)	(0.075)	(0.062)	(0.064)	(0.063)	(0.055)	(0.038)	(0.035)	(0.037)	(0.035)
Emotions test controls										
$Valence_2 - Valence1$ emotion10	0.010					0.019*				
	(0.014)					(0.010)				
$Arousal_2 - Arousal1$	-0.006					-0.003				
	(0.013)					(0.011)				
Sociodemographic controls										
Age		-0.008					0.011			
		(0.010)					(0.009)			
Female participant		0.040					-0.040			
		(0.042)					(0.043)			
Number of siblings		-0.014					0.010			
		(0.021)					(0.024)			
Number of household members		-0.013					-0.011			
		(0.010)					(0.017)			
Socioeconomic strata		0.010					-0.010			
		(0.021)					(0.016)			
Big Five inventory control										
Personality index			-0.035					0.011		
			(0.053)					(0.042)		
Other controls										
Econ and finance students				-0.004					0.041	
				(0.047)					(0.037)	
Number of automatic tosses by computer					0.041					
					(0.031)					
Constant	0.052	0.222	0.047	0.045	0.044	0.088***	-0.059	0.096***	0.087***	0.096***
	(0.057)	(0.266)	(0.054)	(0.057)	(0.055)	(0.024)	(0.216)	(0.020)	(0.025)	(0.021)
Observations	69	69	69	69	69	70	70	70	70	70
Clusters	69	69	69	69	69	70	70	70	70	70

Notes: OLS regressions. Dependent variable is the change in the normalized RET performance in the first group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference category is no exclusion. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Table B.2: OLS results for within-subject changes in reported valence by exclusion, by role. Detailed table

	$\frac{Valence_2 - Valence_1}{Valence_1}$															
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Potential Offender				*Potential Victim*				*C Bystander*				*D Bystander*			
Less than 3 exclusions by part	0.773**	0.693**	0.687**	0.693**	-1.129***	-1.178***	-1.272***	-1.178***	-0.609	-0.076	-0.069	-0.132	-0.680	-0.711	-0.730	-0.805*
	(0.312)	(0.335)	(0.342)	(0.341)	(0.365)	(0.429)	(0.385)	(0.407)	(0.428)	(0.457)	(0.444)	(0.398)	(0.524)	(0.429)	(0.443)	(0.456)
3 or more exclusions by part	0.255	0.205	0.175	0.356	-2.968***	-3.079***	-3.023***	-3.057***	-0.745	-0.615	-0.613	-0.678	-0.539	-0.579	-0.527	-0.602
	(0.400)	(0.432)	(0.402)	(0.403)	(0.567)	(0.556)	(0.514)	(0.523)	(0.460)	(0.533)	(0.536)	(0.497)	(0.556)	(0.510)	(0.533)	(0.527)
Sociodemographic controls																
Age	-0.064				-0.155				0.008				-0.009			
	(0.080)				(0.101)				(0.082)				(0.051)			
Female participant	-0.082				0.405				0.203				-0.139			
	(0.335)				(0.510)				(0.366)				(0.311)			
Number of siblings	0.043				0.446				-0.074				0.259			
	(0.138)				(0.306)				(0.147)				(0.177)			
Number of household members	-0.170				-0.139				-0.316*				0.102			
	(0.121)				(0.263)				(0.172)				(0.157)			
Socioeconomic strata	-0.131				-0.163				0.418***				-0.187			
	(0.172)				(0.188)				(0.141)				(0.152)			
Big Five inventory control																
Personality index		-0.241				0.554				-0.043				0.601		
		(0.499)				(0.717)				(0.577)				(0.485)		
Other controls																
Econ and finance students			0.353				1.212***				0.005				-0.329	
			(0.398)				(0.388)				(0.381)				(0.315)	
Number of automatic tosses by computer by part				-3.800***								-0.821				1.138***
				(0.285)								(0.673)				(0.157)
Number of automatic tosses by computer by part = o,								-								
Constant	1.857	-0.536*	-0.634**	-0.556*	3.953	0.433	0.175	0.444	-0.395	0.002	-0.001	0.091	0.802	0.648	0.740*	0.667
	(2.076)	(0.279)	(0.306)	(0.286)	(2.534)	(0.320)	(0.296)	(0.283)	(1.853)	(0.400)	(0.413)	(0.350)	(1.794)	(0.410)	(0.433)	(0.428)
Observations	69	69	69	69	70	70	70	70	69	69	69	69	70	70	70	70
Clusters	69	69	69	69	70	70	70	70	69	69	69	69	70	70	70	70

Notes: OLS regressions. Dependent variable is the change in the reported valence dimension of emotions in the first group stage from Part 2 compared to the baseline. Reference category is no exclusion. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Table B.3: OLS results for within-subject changes in reported arousal by exclusion, by role. Detailed table

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	$\frac{Arousal_2 - Arousal_1}{Arousal_1}$															
	Potential Offender				*Potential Victim*				*C Bystander*				*D Bystander*			
Less than 3 exclusions by part	-0.345	-0.307	-0.316	-0.307	-0.231	-0.146	-0.203	-0.144	-0.750	-0.714	-0.572	-0.609	0.223	0.231	0.285	0.295
	(0.532)	(0.476)	(0.498)	(0.485)	(0.407)	(0.401)	(0.338)	(0.369)	(0.593)	(0.538)	(0.527)	(0.541)	(0.402)	(0.362)	(0.375)	(0.381)
3 or more exclusions by part	0.087	0.192	0.053	0.156	-0.057	0.255	0.330	0.308	-0.907	-1.028	-0.935	-0.985	-0.000	0.016	0.051	0.036
	(0.486)	(0.495)	(0.500)	(0.504)	(0.584)	(0.499)	(0.431)	(0.461)	(0.691)	(0.652)	(0.646)	(0.652)	(0.355)	(0.359)	(0.342)	(0.343)
Sociodemographic controls																
Age	-0.130				0.015				-0.017				0.026			
	(0.127)				(0.071)				(0.096)				(0.046)			
Female participant	-0.111				0.489				0.775				-0.919*			
	(0.360)				(0.450)				(0.513)				(0.472)			
Number of siblings	-0.390***				-0.210				-0.039				-0.224			
	(0.118)				(0.239)				(0.252)				(0.158)			
Number of household members	0.038				0.331				-0.193				0.067			
	(0.089)				(0.224)				(0.209)				(0.189)			
Socioeconomic strata	-0.127				-0.020				0.054				-0.024			
	(0.169)				(0.177)				(0.175)				(0.202)			
Big Five inventory control																
Personality index		0.478				1.385*				-0.600				-0.506		
		(0.505)				(0.699)				(0.617)				(0.644)		
Other controls																
Econ and finance students			0.505				0.753*				-0.843**				-0.067	
			(0.448)				(0.398)				(0.387)				(0.426)	
Number of automatic tosses by computer by part				-0.600*								0.098				-0.517
				(0.306)								(0.874)				(0.354)
Number of automatic tosses by computer by part = o,								-								
Constant	4.186	0.405	0.332	0.444	-1.161	0.083	-0.056	0.111	0.813	0.471	0.632	0.434	0.490	0.238	0.237	0.222
	(3.108)	(0.391)	(0.439)	(0.400)	(1.786)	(0.316)	(0.205)	(0.251)	(2.359)	(0.511)	(0.533)	(0.521)	(1.831)	(0.178)	(0.172)	(0.143)
Observations	69	69	69	69	70	70	70	70	69	69	69	69	70	70	70	70
Clusters	69	69	69	69	70	70	70	70	69	69	69	69	70	70	70	70

Notes: OLS regressions. Dependent variable is the change in the reported arousal dimension of emotions in the first group stage from Part 2 compared to the baseline. Reference category is no exclusion. Standard errors clustered at the group level. *** p<0.01, ** p<0.05, * p<0.1.

Table B.4: Punishment behavior of bystanders, with and without role reversal

	Sum of opinions of C and D by round		Sum of opinions of C and D by round (RR)	
	(1)	(2)	(3)	(4)
Ostracism in the round	-7.049***	-7.061***	-5.943***	-5.782***
	(0.801)	(0.837)	(0.959)	(0.972)
Age		0.259**		0.0788
		(0.102)		(0.175)
Female participant		-0.193		0.554
		(0.582)		(0.721)
Number of siblings		-0.171		-0.109
		(0.252)		(0.339)
Number of household members		0.189		0.687**
		(0.238)		(0.329)
Socioeconomic strata		0.355*		0.518**
		(0.202)		(0.222)
Personality, using BF		-0.498		-1.510*
		(0.989)		(0.879)
Econ and finance students		-0.627		-1.421*
		(0.471)		(0.717)
Number of automatic tosses by computer by part		-0.792		3.223
		(0.776)		(2.482)
Constant	5.492***	-1.389	3.858***	-1.951
	(0.431)	(2.893)	(0.534)	(4.050)
Observations	695	695	695	695
Clusters	35	35	35	35

Notes: OLS regressions. Dependent variable: sum of opinions of both bystanders, by round. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Table B.5: OLS results for within-subject changes in RET performance by punishment, for bystanders

	$\frac{RET_3-RET_1}{RET_1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	C Bystander				*D Bystander*			
Number of punishment events by part	-0.014	-0.033			0.010	0.034		
	(0.045)	(0.059)			(0.044)	(0.033)		
Less than 3 exclusions by part	-0.049	-0.010	-0.099	-0.059	0.062	0.083	0.072*	0.118
	(0.080)	(0.095)	(0.061)	(0.100)	(0.069)	(0.068)	(0.041)	(0.098)
3 or more exclusions by part	-0.056	0.037	-0.213*	-0.139	-0.011	-0.004	0.108	0.157
	(0.073)	(0.083)	(0.105)	(0.148)	(0.080)	(0.067)	(0.074)	(0.099)
Punishment events × Less than 3 exclusions	0.012	0.022			-0.016	-0.061		
	(0.051)	(0.073)			(0.050)	(0.041)		
Punishment events × 3 or more exclusions	0.000	0.009			0.007	-0.028		
	(0.046)	(0.061)			(0.047)	(0.034)		
Punished once or twice			-0.043	-0.078			-0.026	0.009
			(0.076)	(0.103)			(0.088)	(0.074)
Punished more than twice			0.093	0.068			-0.018	-0.120
			(0.075)	(0.114)			(0.059)	(0.089)
Punished once or twice × Less than 3 exclusions			0.070	0.082			-0.007	-0.093
			(0.079)	(0.130)			(0.087)	(0.110)
Punished once or twice × 3 or more exclusions			0.198*	0.229			-0.121	-0.186
			(0.117)	(0.214)			(0.136)	(0.130)
Constant	0.097	0.311	0.099	0.213	0.029	0.461*	0.033	0.480
	(0.060)	(0.362)	(0.061)	(0.350)	(0.040)	(0.250)	(0.040)	(0.286)
Observations	68	68	68	68	70	70	70	70
Clusters	35	35	35	35	35	35	35	35
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: OLS regressions. Dependent variable is the change in the RET performance on the second group stage from Part 2 compared to the baseline, normalized by the result during the baseline task to account for ability. Reference categories in columns (3), (4), (7), and (8) are no exclusion and no punishment. Controls: changes in emotions (between the second group stage of Part 2 and the baseline), sociodemographic variables (age, female, number of siblings, number of household members, socioeconomic status), personality traits from the Big Five Inventory, economics or finance major, and a measure of inattention. Standard errors clustered at the super-group level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix C

Chapter 3

Figure C.1: Experimental screen: encryption task

Table C.1: OLS regressions of decision to select a tournament compensation, with and without externality. Top performing participants

	Stage 2			Stage 3			Stages 2 and 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.298*	-0.286	-0.223	-0.422***	-0.344**	-0.280*	-0.360***	-0.315**	-0.252*
	(0.162)	(0.170)	(0.171)	(0.150)	(0.146)	(0.153)	(0.122)	(0.127)	(0.128)
Correct encryptions during task	0.110	0.104	0.136	0.063	0.027	0.060	0.086	0.066	0.098
	(0.145)	(0.153)	(0.138)	(0.099)	(0.100)	(0.088)	(0.096)	(0.106)	(0.091)
Age	0.032	0.030	0.045	0.049	0.033	0.048	0.041	0.032	0.046
	(0.041)	(0.042)	(0.042)	(0.035)	(0.043)	(0.041)	(0.033)	(0.037)	(0.035)
Number of female siblings	-0.198***	-0.204***	-0.206***	-0.086	-0.125	-0.127	-0.142**	-0.165***	-0.167***
	(0.069)	(0.069)	(0.069)	(0.081)	(0.079)	(0.080)	(0.060)	(0.061)	(0.060)
Number of male siblings	0.017	0.007	0.003	-0.055	-0.119	-0.124	-0.019	-0.056	-0.061
	(0.123)	(0.129)	(0.129)	(0.120)	(0.110)	(0.114)	(0.090)	(0.090)	(0.092)
Economic status	-0.009	-0.017	-0.051	0.060	0.011	-0.024	0.025	-0.003	-0.037
	(0.096)	(0.102)	(0.086)	(0.095)	(0.093)	(0.082)	(0.083)	(0.084)	(0.069)
Mother's schooling above father's	0.208	0.219	0.212	0.056	0.130	0.123	0.132	0.174	0.167
	(0.148)	(0.142)	(0.144)	(0.141)	(0.129)	(0.128)	(0.113)	(0.106)	(0.105)
Economics or finance student	0.214	0.187	0.248	0.326**	0.151	0.213	0.270**	0.169	0.230**
	(0.160)	(0.199)	(0.189)	(0.152)	(0.151)	(0.130)	(0.111)	(0.132)	(0.110)
Preference for equal payment conditions		-0.074	-0.043		-0.467***	-0.435**		-0.271*	-0.239
		(0.221)	(0.221)		(0.160)	(0.159)		(0.147)	(0.144)
Willingness to take risks			0.146*			0.147**			0.146**
			(0.081)			(0.070)			(0.065)
Externality							-0.073	-0.073	-0.073
							(0.078)	(0.078)	(0.079)
Constant	-1.178	-1.002	-2.050	-1.394	-0.285	-1.339	-1.250	-0.607	-1.658
	(2.060)	(2.221)	(2.187)	(1.485)	(1.523)	(1.540)	(1.347)	(1.494)	(1.473)
Observations	41	41	41	41	41	41	82	82	82
Clusters	-	-	-	-	-	-	82	82	82

Notes: OLS regressions for top performing participants. Dependent variable is the preference for tournament payment. Columns (1) to (3) include the decisions without externality (Stage 2); Columns (4) to (6) include the decisions with externality (Stage 3). Columns (7) to (9) include the decisions in both stages, and Externality is a dummy variable that takes the value of 1 for Stage 3. Robust standard errors in parenthesis in Columns (1) to (6); Standard errors clustered at the individual level (clusters=82) in Columns (7) to (9). *** p<0.01, ** p<0.05, * p<0.1.

Table C.2: OLS regressions of decision to select a tournament compensation, with and without externality. Belief of top position.

	Stage 2				Stage 3				Stages 2 and 3			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.016	0.004	0.025	0.035	-0.275**	-0.102	-0.223	-0.076	-0.145	-0.049	-0.099	-0.020
	(0.134)	(0.158)	(0.138)	(0.157)	(0.132)	(0.127)	(0.134)	(0.126)	(0.096)	(0.103)	(0.099)	(0.104)
Correct encryptions during task	0.008	0.007	0.012	0.011	0.011	0.001	0.016	0.005	0.009	0.004	0.014	0.008
	(0.044)	(0.045)	(0.042)	(0.043)	(0.051)	(0.040)	(0.047)	(0.038)	(0.033)	(0.032)	(0.031)	(0.031)
Age	0.053*	0.056*	0.054*	0.056*	0.006	0.036	0.007	0.035	0.029	0.046**	0.030	0.046**
	(0.030)	(0.029)	(0.028)	(0.028)	(0.031)	(0.025)	(0.031)	(0.026)	(0.024)	(0.021)	(0.023)	(0.021)
Number of female siblings	0.009	0.003	0.019	0.016	0.025	-0.021	0.038	-0.011	0.017	-0.009	0.028	0.002
	(0.079)	(0.081)	(0.083)	(0.085)	(0.086)	(0.067)	(0.090)	(0.071)	(0.058)	(0.055)	(0.060)	(0.058)
Number of male siblings	-0.020	-0.028	-0.035	-0.039	-0.152*	-0.219***	-0.171**	-0.228***	-0.086	-0.123**	-0.103*	-0.134**
	(0.078)	(0.078)	(0.076)	(0.079)	(0.080)	(0.065)	(0.078)	(0.066)	(0.057)	(0.054)	(0.056)	(0.054)
Economic status	0.019	0.015	-0.004	-0.006	0.108*	0.074	0.079	0.057	0.063	0.044	0.038	0.025
	(0.062)	(0.064)	(0.065)	(0.067)	(0.062)	(0.052)	(0.062)	(0.055)	(0.044)	(0.043)	(0.045)	(0.044)
Mother's schooling above father's	0.151	0.167	0.154	0.162	-0.148	-0.011	-0.145	-0.015	0.001	0.078	0.004	0.074
	(0.119)	(0.129)	(0.123)	(0.129)	(0.133)	(0.115)	(0.131)	(0.114)	(0.088)	(0.086)	(0.089)	(0.086)
Economics or finance student	0.250*	0.241*	0.275**	0.270*	0.074	0.001	0.105	0.025	0.162	0.121	0.190*	0.147
	(0.130)	(0.137)	(0.129)	(0.135)	(0.154)	(0.120)	(0.153)	(0.119)	(0.102)	(0.099)	(0.102)	(0.100)
Preference for equal payment conditions		-0.069		-0.039		-0.587***		-0.562***		-0.328***		-0.300***
		(0.180)		(0.176)		(0.112)		(0.121)		(0.105)		(0.107)
Willingness to take risks			0.083	0.081			0.103*	0.068			0.093**	0.074*
			(0.067)	(0.067)			(0.057)	(0.050)			(0.045)	(0.042)
Externality									-0.254***	-0.254***	-0.254***	-0.254***
									(0.087)	(0.083)	(0.085)	(0.082)
Constant	-0.729	-0.757	-1.011	-1.018	0.056	-0.188	-0.295	-0.406	-0.209	-0.346	-0.525	-0.585
	(0.880)	(0.856)	(0.810)	(0.803)	(0.968)	(0.800)	(0.955)	(0.795)	(0.682)	(0.614)	(0.657)	(0.600)
Observations	59	59	59	59	59	59	59	59	118	118	118	118
Clusters	-	-	-	-	-	-	-	-	59	59	59	59

Notes: OLS regressions for those who believe they are top participants in the group. Dependent variable is the preference for tournament payment. Columns (1) to (4) include the decisions without externality (Stage 2); Columns (5) to (8) include the decisions with externality (Stage 3). Columns (9) to (12) include the decisions in both stages, and Externality is a dummy variable that takes the value of 1 for Stage 3. Robust standard errors in parenthesis in Columns (1) to (8); Standard errors clustered at the individual level (clusters=59) in Columns (9) and (12). *** p<0.01, ** p<0.05, * p<0.1.

Table C.3: OLS regressions of decision to select a tournament compensation with information on group composition

	Stage 5				Stages 4 and 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.237***	-0.180***	-0.126**	-0.074	-0.181**	-0.129*	-0.060
	(0.069)	(0.064)	(0.060)	(0.062)	(0.072)	(0.068)	(0.069)
Correct encryptions during task	0.039*	0.032*	0.026	0.023	0.029	0.023	0.019
	(0.020)	(0.019)	(0.017)	(0.017)	(0.018)	(0.017)	(0.017)
Age	0.008	0.005	0.003	-0.003	0.008	0.006	-0.001
	(0.017)	(0.017)	(0.017)	(0.016)	(0.013)	(0.013)	(0.013)
Number of female siblings	0.070*	0.055	0.072**	0.066*	0.021	0.007	0.001
	(0.041)	(0.036)	(0.034)	(0.034)	(0.036)	(0.031)	(0.029)
Number of male siblings	0.033	0.028	0.022	0.019	0.026	0.022	0.018
	(0.045)	(0.041)	(0.036)	(0.037)	(0.040)	(0.038)	(0.037)
Economic status	-0.080**	-0.086***	-0.076***	-0.086***	-0.049*	-0.054**	-0.065**
	(0.032)	(0.029)	(0.026)	(0.026)	(0.029)	(0.027)	(0.026)
Mother's schooling above father's	-0.089	-0.046	-0.039	-0.040	-0.069	-0.031	-0.032
	(0.069)	(0.064)	(0.056)	(0.056)	(0.062)	(0.057)	(0.056)
Economics or finance student	0.083	0.004	-0.016	0.001	0.097	0.026	0.046
	(0.076)	(0.072)	(0.060)	(0.060)	(0.069)	(0.065)	(0.064)
All men groups	0.073	0.122	0.088	0.138	0.058	0.102	0.160
	(0.295)	(0.385)	(0.245)	(0.290)	(0.283)	(0.362)	(0.404)
Preference for equal payment conditions		-0.378***	-0.250***	-0.227***		-0.341***	-0.301***
		(0.068)	(0.068)	(0.068)		(0.059)	(0.060)
Decision to participate in tournament in stage 4			0.421***	0.391***			
			(0.069)	(0.072)			
Willingness to take risks				0.073***			0.090***
				(0.026)			(0.025)
Info on group gender composition					0.000	0.000	0.000
					(0.045)	(0.046)	(0.046)
Female × Info on group composition					-0.049	-0.049	-0.049
					(0.068)	(0.068)	(0.069)
Constant	0.280	0.555	0.353	0.290	0.270	0.518	0.423
	(0.434)	(0.457)	(0.432)	(0.418)	(0.372)	(0.362)	(0.358)
Observations	203	203	203	203	406	406	406
Clusters	-	-	-	-	203	203	203

Notes: OLS regressions. Dependent variable is the preference for tournament payment in groups with information of gender composition. Columns (1) to (4) include the individual preferences for group's tournament with information of gender composition (Stage 5); Columns (5) to (7) include the preferences for group's tournament without and with information on group composition (Stages 4 and 5). Info on group gender composition is a dummy variable that takes the value of 1 for Stage 5. Robust standard errors in parenthesis in Columns (1) to (4); Standard errors clustered at the individual level (clusters=203) in Columns (5) to (7). *** p<0.01, ** p<0.05, * p<0.1.

Table C.4: OLS regressions of decision to select a tournament compensation with information on group composition. Top performing participants

	Stage 5				Stages 4 and 5		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.352**	-0.260*	-0.316**	-0.242*	0.087	0.198	0.256*
	(0.158)	(0.152)	(0.145)	(0.140)	(0.176)	(0.145)	(0.145)
Correct encryptions during task	-0.008	-0.050	-0.027	-0.007	-0.017	-0.067	-0.037
	(0.121)	(0.096)	(0.086)	(0.085)	(0.114)	(0.073)	(0.077)
Age	-0.035	-0.054	-0.038	-0.031	-0.034	-0.056	-0.043
	(0.054)	(0.054)	(0.044)	(0.043)	(0.053)	(0.046)	(0.039)
Number of female siblings	0.050	0.004	0.021	0.014	0.024	-0.031	-0.033
	(0.097)	(0.091)	(0.085)	(0.087)	(0.079)	(0.067)	(0.059)
Number of male siblings	0.131	0.056	0.050	0.048	0.129	0.039	0.035
	(0.105)	(0.087)	(0.094)	(0.096)	(0.100)	(0.070)	(0.066)
Economic status	-0.117	-0.175**	-0.144*	-0.184**	-0.078	-0.146**	-0.178***
	(0.094)	(0.078)	(0.082)	(0.078)	(0.080)	(0.061)	(0.061)
Mother's schooling above father's	-0.225	-0.138	-0.126	-0.136	-0.196	-0.092	-0.098
	(0.163)	(0.152)	(0.146)	(0.145)	(0.144)	(0.122)	(0.117)
Economics or finance student	0.086	-0.120	-0.057	-0.025	0.068	-0.179	-0.122
	(0.217)	(0.196)	(0.169)	(0.167)	(0.185)	(0.158)	(0.139)
Preference for equal payment conditions		-0.550***	-0.349	-0.390*		-0.659***	-0.630***
		(0.168)	(0.220)	(0.203)		(0.116)	(0.105)
Decision to participate in tournament in stage 4			0.262	0.173			
			(0.175)	(0.162)			
Willingness to take risks				0.126**			0.135***
				(0.049)			(0.044)
Info on group gender composition					0.208**	0.208**	0.208**
					(0.090)	(0.090)	(0.091)
Female × Info on group composition					-0.444**	-0.444**	-0.444**
					(0.169)	(0.170)	(0.171)
Constant	1.847	3.154	2.271	1.665	1.593	3.157*	2.189
	(2.400)	(2.007)	(1.635)	(1.632)	(2.215)	(1.638)	(1.566)
Observations	41	41	41	41	82	82	82
R-squared					0.192	0.404	0.472
Clusters	-	-	-	-	41	41	41

Notes: OLS regressions for top performing participants. Dependent variable is the preference for tournament payment in groups with information of gender composition. Columns (1) to (4) include the individual preferences for group's tournament with information of gender composition (Stage 5); Columns (5) to (7) include the preferences for group's tournament without and with information on group composition (Stages 4 and 5). Info on group gender composition is a dummy variable that takes the value of 1 for Stage 5. Robust standard errors in parenthesis in Columns (1) to (4); Standard errors clustered at the individual level (clusters=41) in Columns (5) to (7). *** p<0.01, ** p<0.05, * p<0.1.

Table C.5: OLS regressions of decision to select a tournament compensation with information of group composition

	(1)	(2)	(3)	(4)	(5)
	Stages 3 and 5				
Female majority	-0.115**	-0.087	-0.053	-0.027	-0.011
	(0.055)	(0.056)	(0.049)	(0.047)	(0.049)
Information and group interaction	0.207***	0.207***	0.207***	0.207***	0.207***
	(0.034)	(0.034)	(0.034)	(0.035)	(0.035)
Belief group is better that other group in super-group	0.124	0.142*	0.153*	0.134*	0.125
	(0.087)	(0.083)	(0.079)	(0.079)	(0.078)
Correct encryptions during task	0.039**	0.036**	0.029*	0.024*	0.023
	(0.017)	(0.017)	(0.015)	(0.014)	(0.014)
Preference for equal payment conditions			-0.344***	-0.292***	-0.288***
			(0.047)	(0.047)	(0.047)
Willingness to take risks				0.095***	0.089***
				(0.019)	(0.020)
Female					-0.054
					(0.050)
Constant	-0.141	-0.532	-0.260	-0.303	-0.260
	(0.158)	(0.350)	(0.369)	(0.349)	(0.350)
Observations	406	406	406	406	406
Clusters	203	203	203	203	203
Controls	Yes	Yes	Yes	Yes	Yes
Yes					
	Stages 4 and 5				
Female majority	-0.090	-0.062	-0.026	0.001	0.028
	(0.062)	(0.064)	(0.057)	(0.055)	(0.057)
Information and group interaction	-0.025	-0.025	-0.025	-0.025	-0.025
	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Belief group is better that other group in super-group	0.067	0.093	0.106	0.086	0.070
	(0.098)	(0.099)	(0.091)	(0.089)	(0.087)
Correct encryptions during task	0.035*	0.032*	0.024	0.019	0.018
	(0.018)	(0.019)	(0.017)	(0.017)	(0.017)
Preference for equal payment conditions			-0.364***	-0.310***	-0.304***
			(0.059)	(0.060)	(0.060)
Willingness to take risks				0.100***	0.090***
				(0.024)	(0.025)
Female					-0.090
					(0.060)
Constant	0.164	0.065	0.354	0.309	0.381
	(0.172)	(0.380)	(0.367)	(0.365)	(0.364)
Observations	406	406	406	406	406
Clusters	203	203	203	203	203
Controls	Yes	Yes	Yes	Yes	Yes
Yes					

Notes: OLS regressions. Dependent variable is the preference for tournament payment. The table includes decisions in two stages (3 and 5) and (4 and 5). Information and group interaction is a variable that takes the value of 1 for Stage 5. Controls: age, number of female and male siblings, economic status, mother's schooling level above that of father's, economics or finance student, and all-men groups. Standard errors clustered at the individual level (clusters=203). *** p<0.01, ** p<0.05, * p<0.1.

Table C.6: Summary statistics and mean comparison tests, by treatment

	Mean		p-value
	Random (N=101)	Merit (N=102)	Random vs Merit
Sociodemographic Characteristics:			
Age	21.26	21.09	0.313
Female	0.45	0.57	0.070
Number of female siblings	0.89	0.82	0.117
Number of male siblings	0.78	0.71	0.112
Strata	3.50	3.64	0.149
Mother's level of education above of father's	0.37	0.36	0.068
University information:			
Economics or finance	0.43	0.37	0.069
RET results:			
Trial result	6.96	6.85	0.244
Trial mistakes	0.22	0.20	0.088
Task result	8.36	8.26	0.218
Task mistakes	0.24	0.15	0.070
Other			
Participation in previous experiments	0.87	0.84	0.049
Preference for equal payment conditions	0.37	0.42	0.069
Willingness to take risks	2.99	2.82	0.168
Agreeableness index	-0.01	0.01	0.023

Notes: This table shows the mean by treatment (random and merit) for the final survey variables and the results in the RET. It also includes the p-value for the mean differences between treatments. I present results for age, the proportion of women, number of siblings by sex, and socioeconomic status following the Colombian classification. I also include a dummy that takes the value of one when the mother's schooling level is above that of the father's and another dummy for Economics or Finance majors. In addition, I present the performance during the encryption task and the frequency of the participation in previous experiments. I also include the percentage of participants choosing Alternative 1 in Stage 1 as a proxy for other-regarding preferences. I build an index factoring the nine questions for the agreeableness dimension of the Big Five personality test and present the results for the question on risk preference.