Essays on Public Goods and Social Learning

by

Mustafa Yahşi

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Economics

KOÇ ÜNİVERSİTESİ

October 2023

Essays on Public Goods and Social Learning

Koç University

Graduate School of Social Sciences and Humanities

This is to certify that I have examined this copy of a doctoral dissertation by

Mustafa Yahşi

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Prof. Seda Ertaç (Advisor)

_____

Prof. Mehmet Yiğit Gürdal (Co-advisor)

_____

Prof. Levent Koçkesen

_____

Assoc. Prof. Muhammed Ali Yıldırım

_____

Assoc. Prof. Orhan Torul

_____

Prof. Ayça Ebru Giritligil

_____

Asst. Prof. Emre Ekinci

Date: _____

Statement of Authorship

This thesis contains no material which has been accepted for any award or degree or diploma in any university or other institution.

Chapter 1 is a joint work with Prof. Mehmet Yiğit Gürdal and Prof. Özgür Gürerk. This chapter has been previously published in *Journal of Behavioral and Experimental Economics* in 2021 with the title "Culture and prevalence of sanctioning institutions".

Chapter 2 is a joint work with Prof. Mehmet Yiğit Gürdal and Assoc. Prof. Orhan Torul. This chapter has been previously revised and resubmitted to *Journal of the Economic Science Association* in 2023 with the title "The Seeds of Success: The Pivotal Role of First Round Cooperation in Public Goods Games".

Chapter 3 is a sole-authorship work and exclusively mine.

_____ _____

Signature Name

ACKNOWLEDGMENTS

First and foremost, I would like to express my indebtedness and special thanks to Professor Mehmet Yiğit Gürdal for his exceptional guidance, invaluable insights, and generous support throughout my Ph.D. study. His guidance, insights, and generous support during my Ph.D. study have been indispensable. I have greatly benefited from his unwavering passion, dedication, and authentic enthusiasm for both research and teaching, and I feel incredibly privileged to have had the opportunity to work alongside him.

I would like to extend my heartfelt appreciation to Professor Seda Ertaç for her guidance. Working with her has been an enriching experience, and her valuable expertise has significantly contributed to every stage of my research.

I would also like to express my gratitude to Associate Professor Orhan Torul and Associate Professor Muhammed Ali Yıldırım for their interest in my research. Their insightful suggestions and inquiries have been instrumental in the development of my work.

I extend my sincere thanks to all members within the Department of Economics at Koç University who have contributed, directly or indirectly, to the development of this thesis. I am especially indebted to Esra Öztürk and Duygu Sili for their keen interest in my research and their invaluable contributions to this work.

I would like to extend my warm thanks to all the stray animals of Sarıyer. Their continual affection has brought me endless comfort and joy.

Finally, I would like to express my warmest gratitude to my partner, Gülce Yüce, for her unwavering support and encouragement.

ABSTRACT

Essays on Public Goods and Social Learning
Mustafa Yahşi
Doctor of Philosophy in Economics
October 2023

This dissertation consists of three independent chapters. The first chapter is co-authored with Professor Doctor Mehmet Yiğit Gürdal and Professor Doctor Özgür Gürerk. The second chapter is co-authored with Professor Doctor Mehmet Yiğit Gürdal and Associate Professor Doctor Orhan Torul. The third chapter is solo-authored.

The first chapter investigates the impact of culture on the acceptance and prevalence of sanctioning institutions, and report the results of controlled lab experiments in two countries with substantial cultural differences: Germany and Turkey. We find, if a sanctioning institution is one of two alternatives that individuals can freely choose in an *endogenous* setting, then it is the clear winner against a non-sanctioning institution, both in Germany and Turkey. Though there are some differences in initial institutional preferences and contributions in both countries, the dynamics of institution choice, the evolution of contributions and sanctioning behavior are remarkably similar. Our results extend the findings of a previous study by Herrmann et al. [2008] that finds less cooperative behaviors in some countries (among them Turkey) if the sanctioning institutions are *exogenous*. We show that in one of those countries, Turkey, in an *endogenous* setting, sanctioning institutions beneficial to societies are adapted and prevail.

The second chapter examines the cooperation and punishment in a public goods game in Istanbul. Unlike prior within-subject designs, we implement a between-subject design with separate no-punishment and punishment conditions. This approach reveals that punishment significantly increases contributions, demonstrating the detrimental effect of prior experience without sanctions. We highlight two critical factors - heterogeneous initial contributions across groups and estimated contribution updating rules based on received sanctions. An agent-based model verifies that

their interaction produces strong persistence. Analyzing related data shows similar patterns emerge in comparable cities, suggesting our findings likely generalize if employing a between-subject design. We conclude that overlooking within-group heterogeneity biases cross-society comparisons and subsequent policy implications.

The third chapter analyzes how the structure of the network and the relevancy of the object of interest affects the decisions of the subjects in a social learning experiment. In their most classical form, social learning experiments include a series of subjects lined up in a linear network. In such a design, each subject receives a private signal to predict the state of the world sequentially. The novelty of the chapter stems from its circular line-up design such that the first mover has to make two decisions. Initially, the first mover makes a decision based on her own private signal and another party subsequently makes a decision. Next, the first mover observes the decision made by the other party in its group and updates its prediction. We ask subjects to predict a randomly chosen target number and find that when the predicted item holds no self relevance for any players in the network, we lack sufficient evidence to suggest that the structure of the line-up influences outcomes. We also ask subjects to predict their own and another party's rank in a Raven Progressive Matrices test [Raven et al., 2003]. When the predicted item holds self-relevance for at least one party in the network, we observe a bias in predictions. In particular, the initial assessment of the subjects reveals that there is an overconfidence with respect to their performance in the test. Even though the bias of the predictions decreases compared to their initial beliefs after receiving an information about their rank, they still overestimate their rank. While subjects overestimate their rank more compared to other party's rank with respect to the earlier predictions, the difference between predicting one's own rank and another party's rank vanishes with later predictions.

ÖZETÇE

Kamu Malları ve Sosyal Öğrenme üzerine Makaleler
Mustafa Yahşi
Ekonomi, Doktora
Ekim 2023

Bu doktora tezi birbirinden bağımsız üç ayrı bölümden oluşmaktadır. Birinci bölüm, Profesör Doktor Mehmet Yiğit Gürdal ve Profesör Doktor Özgür Gürerk ile yapılmış ortak bir çalışmadır. İkinci bölüm, Profesör Doktor Mehmet Yiğit Gürdal ve Doçent Doktor Orhan Torul ile yapılmış ortak bir çalışmadır. Üçüncü bölüm tek yazarlı bir çalışmadır.

İlk bölüm, kültürün yaptırım uygulayan kurumların kabulü ve yaygınlığı üzerindeki etkisini inceleyerek, önemli kültürel farklılıklara sahip iki ülkedeki (Almanya ve Türkiye) kontrollü laboratuvar deneylerinin sonuçlarını raporlamaktadır. Yaptırım uygulayan bir kurum, bireylerin içsel bir ayarda özgürce seçebilecekleri iki alternatiften biriyse, o zaman hem Almanya'da hem de Türkiye'de yaptırım uygulamayan bir kuruma karşı açık ara galip gelmektedir. Her iki ülkedeki ilk kurumsal tercihler ve katkılarda birtakım farklılıklar olsa da kurum seçiminin dinamikleri, katkıların evrimi ve yaptırım davranışı dikkate değer ölçüde benzer olarak görülmektedir. Bu çalışma, yaptırım kurumlarının dışsal ayarlı olması durumunda bazı ülkelerde daha az işbirlikçi davranışlar bulan Herrmann et al. [2008] tarafından yapılan önceki bir çalışmanın bulgularını genişletiyor. Bu ülkelerden biri olan Türkiye'de içsel ayarlı bir ortamda, toplumlara faydalı yaptırım kurumlarının uyarlandığı ve baskın geldiği görülmektedir.

İkinci bölüm, İstanbul'da yapılan bir kamu malı deneyinde iş birliği ve ceza arasındaki ilişkiyi incelemektedir. Önceki denek-içi dizaynlardan farklı olarak, birbirinden ayrı cezasızlık ve ceza koşullarına sahip denekler arası bir dizayn seçimi yapılmıştır. Bu yaklaşım, cezanın katkıları önemli ölçüde artırdığını ve yaptırımlar olmadan önceki deneyimlerin zararlı etkisini ortaya koymaktadır. İki kritik faktörün altını çizmek gerekmektedir: gruplar arasında heterojen başlangıç katkıları ve karşılaşılan yaptırımlara dayalı olarak tahmini katkı güncelleme kuralları. Aracı tabanlı bir

model, bunların etkileşimlerinin güçlü bir kalıcılık ürettiğini doğrulamaktadır. İlgili verilerin analizi, karşılaştırılabilir şehirlerde benzer şablonların ortaya çıktığını gösteriyor; bu da denekler arası bir dizayn kullanıldığında bulgularımızın muhtemelen genelleşebileceğini göstermektedir. Sonuç olarak, grup içi heterojenliği gözden kaçırmanın, toplumlar arası karşılaştırmaları ve müteakip politika çıkarımlarını önyargılı hale getirdiği sonucuna varılmıştır.

Üçüncü bölüm, bir sosyal öğrenme deneyinde ağ yapısının ve tahmin edilen nesnenin deneklerle olan ilişkisinin deneklerin kararlarını nasıl etkilediğini analiz etmektedir. En klasik biçimlerinde, sosyal öğrenme deneyleri, doğrusal bir hatta sıralanan bir dizi deneği içermektedir. Böyle bir dizaynda, her denek, hedef nesneyi tahmin etmek için özel bir sinyal alır. Bu bölümün görece yeniliği, ilk hareket edenin iki karar vereceği dairesel diziliş tasarımından gelmektedir. Başlangıçta, ilk denek kendi özel sinyaline dayalı olarak bir karar verip ardından ikinci denek kendi özel sinyaline bağlı bir karar verir. Daha sonra, ilk denek, grubundaki diğer deneğin verdiği kararı gözlemleyip tahminini günceller. Deneklerden ilk olarak rastgele seçilmiş bir hedef sayıyı tahmin etmelerini istiyoruz. Tahmin edilen ögenin ağdaki herhangi bir denek ile ilgisi olmadığında deneklerin diziliş yapısının sonuçları etkilediğini önermek için yeterli kanıtımız olmadığını görüyoruz. Ardından deneklerden bir şekiller testinde [Raven et al., 2003] kendilerinin ve başka bir deneğin sıralamasını tahmin etmelerini istiyoruz. Tahmin edilen öge, ağdaki en az bir denek ile ilgili olduğunda tahminlerde bir sapma olduğunu gözlemliyoruz. Özellikle deneklerin ilk değerlendirmesi, testteki performanslarına ilişkin aşırı bir güven olduğunu ortaya koymaktadır. Deneklerin kendi sıralamaları hakkında bir bilgi aldıktan sonra tahminlerinin yanlılığı başlangıçtaki inanışlarına göre azalsa da yine de sıralamalarını gerçekte olduğundan daha iyi tahmin etmektedirler. Deneklerin kendi ve diğer deneklerin sıralaması hakkındaki tahminlerini kıyasladığımızda kendi sıralamaları ile ilgili tahminlerde daha fazla sapma olduğu görülmekle beraber bu fark sonraki tahminlerle ortadan kalkmaktadır.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

CULTURE AND PREVALENCE OF SANCTIONING INSTITUTIONS

1.1 Introduction

Differences in social and economic outcomes across countries are often attributed to cultural differences. Guiso et al. [2006] define culture as "those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation." While these beliefs and values are crucial for understanding human behavior and cooperation [Richerson and Boyd, 2005], the extent of the impact of culture on economic prosperity is the subject of an ongoing discussion. Some scholars highlight the predominance of institutions on economic development [Acemoglu and Robinson, 2013]; others argue that culture and institutions co-evolve influencing each other [Bisin and Verdier, 2017].

With empirical (field) data, it is challenging to disentangle both the effects of institutions and the impact of culture on behaviors. There are some few notable exceptions of field studies documenting the role of cultural differences in determining the success of specific institutions. Putnam et al. [1993], for example, report the case of governmental reform in Italy beginning in 1970 that succeeded very differently in different regions. While the adoption of the reforms concerning government performance was a success in North and Central Italian regions, similar institutional arrangements were not that successful in south Italian provinces. The authors explain the unequal impact of this institutional intervention with different cultural trajectories in Northern-Central and Southern Italy, particularly with differences in the historically routed civic engagement.

The experimental methodology can help to gain insights about relationships between culture and institutions that are difficult to observe in the field. In controlled lab settings, we can implement the *same* institutional arrangement in different locations with different cultural trajectories, which enables us to study culture's influence on behavior. Most of the existing experimental studies follow this approach by keeping the institutional arrangement fixed and comparing the outcomes in different cultures.

On the other hand, not much is known about the interaction of culture and institutions when people face a *choice* between two different institutional arrangements. The ability to choose between institutions allows individuals to obtain first-hand experience across multiple institutions, update their choices and behavior, and potentially allow superior institutions to prevail. In this study, we investigate such a situation by utilizing a framework that gives individuals a choice between two communities with different institutional parameters while varying the actors' cultural background. One critical element of our framework is that members of a community can monitor what happens both in their own and the other community.

In our experiment, a fixed group of subjects interact in a modified version of the voluntary contribution mechanism (VCM). A typical VCM environment involves a group of subjects, where each member voluntarily chooses how much to contribute to a non exclusive public good. In the current experiment, subjects are allowed to choose between a community with a VCM institution as well as a community with a sanctioning institution, which is essentially a VCM institution with an additional ability to sanction other group members after observing their contributions. After this choice at each period, two communities emerge, and every subject takes decisions in the chosen community.

Institutions can be defined in different ways, as equilibrium behavior, as norms to people adhere or simply the rules of the game [Crawford and Ostrom, 1995]. Following the frequently used terminology in the recent experimental literature on endogenous choice of institutions [Nicklisch et al., 2016; Fehr and Williams, 2018; Dannenberg and Gallier, 2020], we consider our two communities as two distinct

institutional arrangements, or simply as two institutions.[1]

We find that in Turkey, which exhibits substantial cultural differences compared to Germany, a sanctioning institution wins the *endogenous* institutional competition against a VCM institution, just as it does in Germany. Though the subjects from two cultures exhibit a substantial variation during the initial stages of the experiment, still they converge towards very similar outcomes.

Our results extend the findings of a previous study by Herrmann et al. [2008] that finds less cooperative behaviors in some cultures (among them Turkey) if the sanctioning institutions are *exogenous*. Our results show that in countries with substantially different cultural backgrounds, in an *endogenous* setting, sanctioning institutions beneficial to societies can be adapted and prevail.

1.2 Related literature

Public good experiments, where subjects contribute towards a non-excludable pool, have been extensively studied. Under certain parameter restrictions, the usual design of these experiments induces an *n-person* prisoner's dilemma game where free-riding is the dominant strategy. In repeated plays, usually subjects start with positive contributions, but eventually, free-riding prevails, and contributions exhibit a continuous decline. Sanctioning institutions, however, can successfully resolve the problem of free-riding in public goods environments [Yamagishi, 1986; Ostrom et al., 1992; Fehr and Gächter, 2000]. In addition to this, when given a choice, people overwhelmingly prefer sanctioning institutions over alternatives where this option is not available, and they effectively utilize sanctions to tame free-riding behavior [Gürerk et al., 2006].

Previously conducted public goods experiments reveal striking cross-cultural differences between subject pools from countries like Germany, Switzerland, the USA, and those from Turkey, Greece, and Russia [Herrmann et al., 2008]. There are

[1] As an anonymous reviewer pointed out, one could consider our communities as two subsets of one unique institution where people sort themselves to subgroups, perhaps hoping to meet like-minded others there.

also few studies that examine the cultural differences between different regions of a country.[2] Substantial differences also exist for other contexts, such as bargaining [Roth et al., 1991], truth-telling [Abeler et al., 2019; Gächter and Schulz, 2016], fairness views [Almås et al., 2020], propensity to punish corrupt behavior [Cameron et al., 2009], and betrayal aversion [Bohnet et al., 2008]. Besides, cross-cultural differences in social preferences (assessed via ultimatum and dictator games) are prevalent across small scale societies [Henrich et al., 2005, 2010].

Given the stylized findings on sanctioning in public goods experiments, we ask whether *culture* can influence the prevalence and the 'long-term' success of sanctioning institutions when individuals have a free choice to join and leave institutions. Motivated by this question, this paper compares the findings of a public goods experiment reported in Gürerk et al. [2014], that was run in Germany, with a series of new experiments conducted with undergraduate students in Turkey.

1.2.1 Differences between Germany and Turkey when institutions are exogenous

We think that a comparison between Turkey and Germany provides an ideal setup for answering our main research question, i.e., the impact of culture when sanctioning institutions can be chosen endogenously. We were motivated by the data reported in the seminal paper by Herrmann et al. [2008]. Extensive cross-country data from this study reveals that the sanctioning mechanism is highly effective in some countries, whereas it has no apparent positive effect on contributions in others. According to these results, Germany belongs to the former group and Turkey belongs to the latter.

Herrmann et al. [2008] data also reveal that there is a great deal of heterogeneity in terms of initial contributions. While in Copenhagen (Denmark), St. Gallen (Switzerland), and Boston (USA), the starting contributions tend to be high, in some

[2]Bigoni et al. [2016] and Bigoni et al. [2019] conduct a large scale lab-in-the-field experiment to find that Southern Italians cooperate less than Northerns and attribute this difference to the beliefs about others' cooperative behavior and the preferences toward social risk. On the other hand, Zhang [2018] studies the difference in the propensity to report the corruption between Northern and Southern Italians and observes that exposure to corruption enhances accountability norms merely in the presence of high-quality enforcement institutions.

other cities like Istanbul (Turkey), Riyadh (Saudi Arabia), and Athens (Greece), they are relatively lower. In the first three cities, without sanctioning mechanism, subjects' initial contributions amount to 65-70 percent, while in the latter three subjects contribute only 40-45 percent of their endowments initially. One observes a similar picture also with the sanctioning mechanism: with 75 percent or more, Boston, Copenhagen, and St. Gallen display the highest initial contributions. In Istanbul, Athens, and Riyadh, subjects contribute only 30 percent, even less than without sanctioning mechanism. Concerning overall contributions across both institutions, while Istanbul exhibits the 14th highest average contribution out of 16 cities, Bonn (Germany) has the 7th highest average. Finally, while the sanctioning mechanism leads to a significant improvement in contributions in Germany, it fails to do so in Turkey (see Table 1.1).

Table 1.1: Contributions and payoffs with exogenous institutions

	Contributions in period 1			Contributions over all periods			Payoffs		
	VCM	PUN	*p*-value	VCM	PUN	*p*-value	VCM	PUN	*p*-value
Istanbul - [Gürdal et al., 2020]	8.9	9.1	0.950	6.1	10.9	0.019	23.6	19.4	0.019
Istanbul - [Herrmann et al., 2008]	8.9	6.5	0.034	5.4	7.1	0.326	23.3	17.0	0.001
Bonn - [Herrmann et al., 2008]	10.9	12.1	0.012	9.2	14.5	0.001	25.5	24.1	0.691

Notes: The table reports the average contributions (out of 20 tokens) and the p-values for Mann-Whitney tests that use group average contributions as independent observations and test for the equality of distributions in the sanction free institutions with voluntary contribution mechanism (VCM) and sanctioning institutions with punishment possibilities (PUN).

In a recent study, Gürdal et al. [2020] report the results of a replication of Herrmann et al. [2008] conducted at the very same school in Turkey. Gürdal et al. [2020] find that contributions are significantly higher in the sanctioning institution.[3] However, as Table 1.1 shows, the average contribution in Gürdal et al. [2020] is still lower than those reported for Germany (Bonn) in Herrmann et al. [2008], suggesting

[3]Herrmann et al. [2008] implement a within-subject design where sanctions treatment follows no sanctions treatment while Gürdal et al. [2020] employ a between-subject design, hence subjects either take part in the sanction-free institution or in the sanctioning institution.

robust differences across two cultures. Gürdal et al. [2020] point out the extensive heterogeneity across groups in the presence of the sanctioning institution for relatively lower levels of contribution. Both Gürdal et al. [2020] and Herrmann et al. [2008] state that groups with higher level of contribution in the first period have a comparably higher level of contribution in the subsequent periods.

1.3 Theoretical considerations

The prevalence of cultural differences is often attributed to cultural group selection hypothesis. According to this hypothesis, human beings are endowed with tools to acquire skills, beliefs, attitudes, and values from other individuals [Soltis et al., 1995] and these skills enable individuals to adapt complex social norms. A high level of cultural variation among different cultural groups sets the stage for the evolution of group-beneficial behavior [Richerson et al., 2016]. Henrich [2004] proposes several cultural transmission mechanisms that sustain the variation between different groups. These mechanisms facilitate cultural learning by leading individuals to copy high-frequency behaviors and imitate successful skills. Punishment of norm violators enhances this process further by increasing the fitness of prestigious and norm-abiding individuals, who in turn will be imitated at a higher frequency by the rest of the population.

Weber and Dawes [2005] stress that it is often important what happens before decision-making as the decision itself. They consider the ignorance of trajectories as one incompleteness of traditional economic modeling. In our experiments, participants are able to observe the relative performances of sanctioning and non-sanctioning institutions. Therefore, they make their decisions each period with the information of previous periods. We believe this is a relevant factor that can facilitate the institution choice and convergence in our experiment. Experimental evidence indeed shows that trajectory may affect behavior. Gürerk [2013], for example, reports that if participants of an experiment are informed about the complete and detailed course of a previous run of the same experiment with different subjects, the observed initial behavior changes significantly. In an experiment that uses a similar setup as

in this study, a higher fraction of the *informed* subjects join the sanctioning institution, compared to *uninformed* subjects. Moreover, informed subjects start with higher contributions to the public good.

1.3.1 On the process of endogenous choice: differential migration

Differential migration is one form of intergroup competition [Henrich, 2015]. In particular, this peaceful form of intergroup competition suggests that people from less successful groups may migrate to more thriving societies that display higher cooperation and harmony. Boyd and Richerson [2009] provide a theoretical model on this. One can observe such movements in small-scale societies [Knauft, 1985; Tuzin, 2013], as well as in actual migration patterns from North Africa to Europe, or from Middle and South America to North America. Today, the internet and mobile devices make it easy than ever for individuals from less developed countries to compare his or her personal living conditions with other places.

In the economic domain, a similar argument was made by Tiebout [1956]. Tiebout asserts that open communities can lead to economic efficiencies. If economic agents are free to move, they can migrate to the community offering the level of public goods that suits them most, i.e., they vote with their feet.

In our setting, we could observe a similar differential migration, or a voting-with-feet dynamics as described above. In an endogenous choice setting with two options, namely a sanctioning institution and a sanction-free institution, members of groups that start with lower average contributions in the sanctioning institution may be inclined to switch to the sanction-free institution, while individuals from groups with higher average contributions in the sanction-free institution may be inclined to switch to the sanctioning institution. The dynamics of institutional choice could make endogenous institutions perform differently than corresponding exogenous institutions. Thus, we hypothesize that an endogenous sanctioning institution in Turkey is likely to be as successful in inducing high levels of contributions to the public good as in Germany. In coordination games, it has been shown that slow growth can help small and efficiently coordinated groups to maintain coordination

while the group size gradually increases [Weber, 2006].

1.3.2 On the cultural differences between Germany and Turkey

There are well-documented cultural differences between Germany and Turkey. In this subsection, we will refer to Hofstede [2001] to measure cultural variations across countries, and to Falk et al. [2018] which investigates the country-level differences in terms of economic preferences.

Hofstede [2001] lists power distance, individualism, masculinity, uncertainty avoidance, long term orientation, and indulgence as the six dimensions of culture. The score for each scale ranges between 0 to 100, with 50 being a mid-level. Figure A.1 in the Appendix shows the comparison of all six dimensions. We think that the reported cultural differences between Germany and Turkey in terms of uncertainty avoidance and long term orientation might be relevant in our context. Based on the long term orientation index (Germany: 83 vs. Turkey: 46), one can expect Germans to contribute at relatively higher levels in the early periods in order to sustain high contribution levels in subsequent periods. Similarly, based on uncertainty avoidance index (Germany: 65 vs. Turkey: 85), subjects from Turkey might contribute at lower levels initially to avoid the risk brought about by others' potentially low contributions.

Falk et al. [2018] study the global variation in economic preferences and report cross-country differences for the following: time preference, risk preference, positive and negative reciprocity, altruism, and trust. Each preference measure is standardized at the individual level; hence, each preference has a mean of zero and a standard deviation of one in the individual-level world sample. The respective comparison of Germany and Turkey can be seen in Figure A.2 in the Appendix. In terms of time preferences, Germans are documented to be more patient than the Turks in this survey (0.624 vs. -0.047). In our context, patience might enhance subjects' persistence to contribute relatively higher amounts in order to sustain cooperation in their respective communities. Furthermore, we also observe that Germany scores higher in altruism (-0.051 vs. -0.279), which might also be supportive of expecting higher

initial contributions in Germany.

Cross-country differences, which might have potential effects on cooperative behavior, were also reported in Herrmann et al. [2008]. In the current study, we report the relevant differences between Germany and Turkey in Table A.1 in the Appendix. According to the World Value Survey, Germans trust a larger share of the population (45 percent) than Turks do (12 percent), which might be an explanation for the initial level of contributions. While the current experiment focuses on a setting with endogenous choice of institutions, most of the experimental studies of voluntary contributions mechanism involves an exogenous assignment of a group of subjects to the punishment institution. [Herrmann et al., 2008] have documented important differences in the success of punishment institutions across different cultures and attirubuted these differences to the differences in the extent that the norms of civic cooperation prevail in the society. However, World Value Survey reveal no striking difference across these two countries in terms of norms of civic cooperation, and hence, we expect not to observe significant variation in punishment attributes. Lastly, the difference in the rule of law index across two countries hints that Germans are more inclined to accept the norms/rules of the institution and act accordingly.

1.4 Experimental design and procedures

The experiment is based on the setup used in Gürerk et al. [2006, 2014], therefore we use the same terminology. In the experiment, a group of individuals simultaneously choose between two communities (institutions) before playing a public good game with each other who have chosen the same community. The two communities that subjects could choose among are termed as the VCM community and the PUN community, respectively. While, in the VCM community, subjects merely decide on the contribution level towards the public good; in the PUN community, after being informed about individual contributions, the players may punish others in their community at an additional cost to themselves. The experiment has a voting with feet stage, a voluntary contribution stage and a punishment stage only in the PUN community.

1.4.1 The voting with feet stage

In the voting with feet stage, all members of an independent group choose one of the two communities simultaneously. After all players made their decisions, each player is informed about the number of players who have chosen the same community. A player's identity and her history of the play is concealed from the other players.

1.4.2 The voluntary contribution stage

In this stage, a player i is in a relation only with the players who have chosen the same community. Each player is given 20 monetary units (tokens) and may decide to contribute an integer amount of g_i, with $0 \leq g_i \leq 20$ to a joint project. The players simultaneously choose their own contribution. The amount not contributed stays in the respective player's private account. The sum of all contributions, denoted by G, in the respective community is multiplied by 1.6, and then allocated to each player in this community, irrespective from the player's individual contribution g_i. Consequently, the marginal per capita return depends on the community size n (see Table 1.2). Once all players make their contribution choices, they are informed about the individual contributions of each member in their respective community without the identity of players being revealed.

Table 1.2: Marginal per capita return

Community size n	2	3	4	5	6	7	8	9	10	11	12	13
MPCR	0.80	0.53	0.40	0.32	0.27	0.23	0.20	0.18	0.16	0.15	0.13	0.12

1.4.3 The punishment stage

In the punishment stage, played only by those choosing the PUN community, each player is endowed 20 additional monetary units independent from its own decision in the contribution stage. We provide 20 additional tokens to the players in the VCM community in order to eliminate incentives for the players to choose the punishment

community just to receive extra monetary units. Since there is no punishment in the VCM community, the public good game ends after the contribution stage for those subjects choosing this community. The total monetary payoff for player i in the VCM community is given by

$$\pi_i = (20 - g_i + 1.6G/n) + 20.$$

On the other hand, in the PUN community, all players must simultaneously decide whether to punish other members of their community or not. Each player may allocate up to 20 tokens for punishment. Player i can punish player j of the PUN community by assigning punishment tokens t_{ij}. If player i assigns 1 token to player j, it will reduce the payoff of player i by 1 and the payoff of player j by 3 tokens. Let $\tau_i = \sum_{j \neq i} t_{ij}$ and $\tau_{-i} = \sum_{j \neq i} t_{ji}$. Accordingly, the total monetary payoff of player i in the punishment community is given by

$$\pi_i = (20 - g_i + 1.6G/n) + (20 - \tau_i - 3\tau_{-i})$$

The expressions in parentheses denote the stage payoffs from the contribution stage and the punishment stage, respectively. Once the punishment stage is completed, we inform all the players in the PUN community about contributions, sent punishment tokens to others, received punishment tokens and the resulting total payoffs for all players in that community. We also inform the members of both communities, the VCM and the PUN community, about the individual contributions and payoffs in both communities. With the availability of such open information flow between both communities, subjects may decide to change their community in the subsequent periods. The above sequence of interactions is repeated for consecutive 20 periods during given session.

1.4.4 Experimental procedures

The experiment was conducted with 85 subjects in 2016 at the Economics Laboratory of the Boğaziçi University; 6 of the sessions had each 12 subjects, the remaining

one session had 13 subjects. An e-mail was sent to subjects who previously indicated an interest in joining economics experiments, and subjects could register online for a date and time they choose. Subjects were informed about the experimental procedure upon their arrival, and the instructions were read aloud. No subject participated more than once, and the sessions lasted on average 50 minutes. Subjects were paid in cash at the end of the experiment. We did not pay the subjects any show-up fee. Instead, we provide them with a starting capital of 1000 tokens to prevent bankruptcy due to excessive punishment expenditures. Random reshuffling of the presentation order on the computer screens ensured that the identity of the players could not be traced over periods.

The current experiment uses the same software code based on z-Tree [Fischbacher, 2007], and instructions as the treatment PUN reported in [Gürerk et al., 2014] that was conducted at the University of Erfurt, Germany, in 2003.[4] The experiments in Germany and Turkey share the same design and procedures except the following two differences. First, in Gürerk et al. [2014], the experiment lasts for 30 periods, where this duration was 20 periods in our case. The reason for this was that the original experiments lasted more than 2 hours. Participants are not used to such a long duration for experiments at Boğaziçi University. In addition, from previous studies we know that much of the dynamics happen in the first 10-15 periods. Thus, we decided to shorten the number of rounds to attract a sufficient number of participants. Second, in Turkey, one of the observations has 13 players. This has no effect on the constant productivity of the public good. There is a marginal decrease in MPCR from 0.13 to 0.12, if a community reaches the maximum capacity with 13 individuals instead of 12.

[4]Since quite some time has passed between the two experiments in Germany and Turkey, one may doubt the comparability of the two samples. However, in a series of experiments that were conducted in fall 2015, an independent group of researchers replicated the findings of the original PUN treatment in Germany from 2003 remarkably closely [Chugunova et al., 2020]. Thus, we believe that the time span between between the two treatments in our study is not a relevant issue.

1.5 Results

We start the results section by looking at contribution and payoff differences between Turkey and Germany. While section 1.5.1 involves our main hypotheses, in the remaining sections, we report our rather exploratory results. Section 1.5.2 looks at first period behavior, section 1.5.3 on punishment behavior, and section 1.5.4 focuses on institution choice and change behavior.

The reported non-parametric statistical tests are two tailed and take total population, or the respective communities as units of observations. Accordingly, we treat a group of interacting subjects as one independent observation. This means, we have 7 independent observations in Turkey: six different groups of 12 subjects each and one group with 13 subjects, in total 85 individuals. In Germany, we had 8 independent observations: 8 different groups of subjects, a total of 96 individuals. While comparisons within a treatment are tested with the Wilcoxon matched-pairs tests, comparisons across treatments are conducted with the Mann-Whitney U-tests. In regression analyses, the unit of observation is a single individual. Standard errors are clustered across groups to control for group effects. Accordingly, with respect to the regression analyses we have a total of 85 observations in Turkey and 96 in Germany.

1.5.1 Overall Contributions and Payoffs

First we report some descriptive statistics. Table 1.3 summarizes the mean values and the standard errors of contribution decisions and realized payoffs, separately for the VCM and the PUN communities in Turkey (present study), and those conducted in Germany by Gürerk et al. [2014]. Compared to PUN, the average contribution levels are quite low under VCM, and the standard errors of the contributions are arguably lower as well. In the PUN community, the mean contribution start at levels slightly above 10 and keep increasing for the subsequent periods. In the VCM community, we observe the opposite as the contributions approach to zero towards the end of the experiment. The average payoff attains values close to 40

and the variations is highly limited for the VCM community. This situation can be explained by the stable and low levels of average contributions. We also observe that the standard error of payoff is almost negligible. For the PUN community, compared to VCM, the payoff takes low values in the beginning of the experiment, but keeps increasing for the subsequent periods.

Table 1.3: Contributions and payoffs over time

	Average Contribution				Average Payoff			
	VCM		PUN		VCM		PUN	
	TUR	GER	TUR	GER	TUR	GER	TUR	GER
Period 1	3.14	7.39	10.05	13.07	41.88	44.44	21.25	23.04
	(0.72)	(0.69)	(1.38)	(0.77)	(0.43)	(0.41)	(3.28)	(6.45)
Period 1-10	2.46	3.55	16.88	16.53	41.48	42.13	39.57	35.58
	(0.39)	(0.59)	(0.85)	(0.85)	(0.23)	(0.36)	(2.50)	(2.80)
Period 11-20	1.21	2.07	19.54	19.19	40.73	41.24	49.14	47.95
	(0.83)	(0.55)	(0.24)	(0.40)	(0.50)	(0.33)	(0.86)	(1.33)
Period 21-30	-	1.06	-	19.67	-	39.63	-	48.65
	-	(0.77)	-	(0.11)	-	(1.19)	-	(0.93)
All Periods	2.23	2.92	18.48	18.83	41.34	41.75	45.33	45.64
	(0.52)	(0.57)	(0.43)	(0.31)	(0.31)	(0.34)	(1.23)	(1.13)

TUR and GER stand for the subjects in Turkey and Germany, respectively.

The unit observation is a subject's own contribution to the public good for the mean statistic.

Standard errors are clustered across groups.

In the VCM community, over all periods, the average contribution to the public good is 2.23 for Turkey, while it is 2.92 for Germany ($p = 0.203$). In the PUN community, the average contribution over all periods is 18.48 for Turkey and 18.83 for Germany ($p = 0.643$).

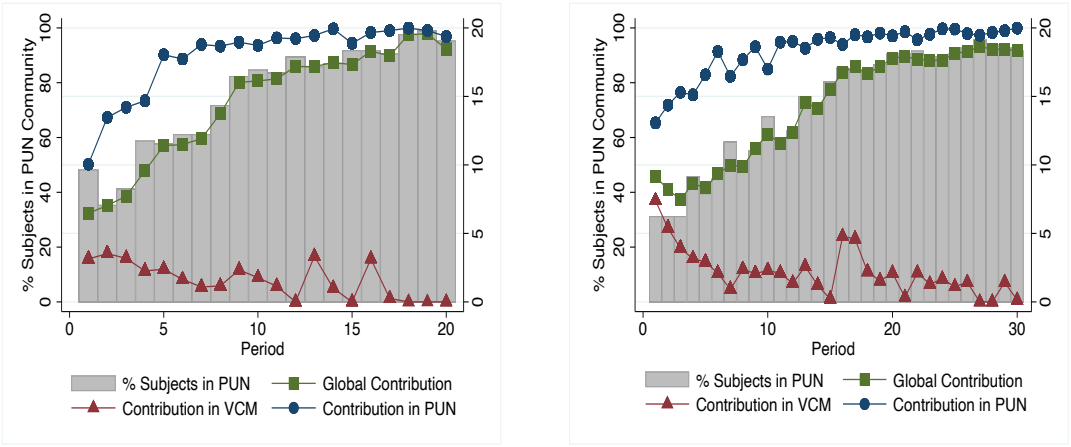
A comparison of the payoff levels reveals no significant difference across countries for both communities. In particular, in the VCM community, the payoff over all periods is on average 41.34 for Turkey, whereas it is 41.75 for Germany ($p = 0.203$). In the PUN community, the average payoff over all periods is 45.33 for Turkey and

45.64 for Germany ($p = 1.000$).

An alternative presentation of the overall results comparing the first and the second half of the respective experiment can be found in the Appendix in Table A.2. It shows that there is no significant difference in the average contributions in the first and second half of the game for the VCM and the PUN communities across countries.

Evolution of Contributions

Figure 1.1 shows the evolution of contributions in Turkey and Germany for each period. In both samples, contributions remain much lower under VCM community compared to the PUN community. Number of subjects in the PUN community gradually increase in both samples, with almost nobody remaining in the VCM community towards the end of the game. For early periods, the percentage of subjects in the PUN community is generally higher in Turkey compared to Germany. Figure A.3 in the Appendix depicts the evolution of contributions for each of the independent groups.



(a) Turkey (b) Germany

Figure 1.1: Evolution of Contributions

Evolution of payoffs by communities

Figure 1.2 shows the evolution of payoffs by communities in Turkey and Germany, respectively. In the VCM community, the payoffs vary less across periods for Turkey than for Germany. Furthermore, payoffs are more volatile in the PUN community in the first periods for Germany. Overall, the evolution of payoffs in both communities follow a similar pattern for Turkey and Germany.

In the Appendix, we provide additional information on the dynamics of payoffs. Figure A.17 shows the payoffs of different types of players and Figure A.18 shows the average payoffs for VCM and PUN communities. Qualitatively, the dynamics of payoffs are very similar in Turkey and Germany, with the average payoffs in the PUN community getting ahead of the VCM community slightly earlier in Turkey.

Result 1 *Over all periods, contributions and payoffs in Turkey and Germany are very similar, both in the VCM and PUN communities. Also, the evolution of contributions and payoffs in both countries are similar.*

(a) Turkey (b) Germany

Figure 1.2: Evolution of payoffs by communities

1.5.2 First period behavior

In this section, we compare the first period behavior of subjects from Turkey with those from Germany. For this comparison, the data indicate some significant differences in contributions and payoffs, however, no significant differences in the initial institutional taste and punishment behavior.

In the VCM community, the number of subjects is on average 6.24 (52%) for Turkey, whereas it is 8.25 (69%) for Germany. In the PUN community, the number of subjects is on average 5.92 (48%) for Turkey, whereas it is 3.75 (31%) for Germany. In the VCM community, the first period contribution to the public good is on average 3.14 for Turkey, whereas this value is on average 7.39 for Germany ($p = 0.003$). In the PUN community, the initial contribution is on average 10.05 for Turkey and 13.07 for Germany ($p = 0.073$), i.e., subjects in Turkey start with lower contributions in both communities.

A comparison of the payoff levels for two samples point out a difference in the VCM communities but not in the PUN communities. In particular, in the VCM community, the first period payoff is on average 41.88 for Turkey, whereas it is 44.44 for Germany ($p = 0.003$). This is not surprising given the higher contributions in Germany for the initial periods. On the other hand, in the PUN community, the first period payoff is on average 21.25 for Turkey and 23.04 for Germany ($p = 0.565$). In PUN community, while subjects assign 0.98 sanction points on average to other subjects in Turkey, it is 1.52 in Germany($p = 0.949$).

Result 2 *Initial preferences for both communities are similar in Turkey and Germany. In the VCM community, contributions in Germany are considerably higher, which leads to higher payoffs. In the PUN communities, contributions in Germany are only weakly significantly higher than contributions in Turkey, while neither average punishment nor payoffs differ significantly between Germany and Turkey.*

Classification of subjects according to first period behavior

To see whether there are some differences between types of subjects, we classify subjects who contribute at least 15 tokens to the public good in the first period but do not punish other subjects as cooperators. We label participants who are cooperators and also punish others as strong cooperators. Finally, we call subjects who contribute 5 or less tokens free-riders. Accordingly, Table 1.4 indicates the mean level of contribution made by cooperators, strong cooperators and free-riders respectively.

Table 1.4: Contribution levels by different types of subjects

	Free-rider		Cooperator		Strong cooperator	
Statistic	TUR	GER	TUR	GER	TUR	GER
Mean	14.52	11.53	-	19.17	18.95	18.44
SE	(0.63)	(1.48)	(-)	(0.24)	(0.16)	(0.53)

TUR stands for the subjects in the respective category (e.g. free riders in Turkey)

GER stands for the subjects in such category (e.g. cooperators in Germany)

Averaged over all periods, free-riders contribute 14.52 in Turkey, while they contribute 11.53 in Germany ($p = 0.197$). Moreover, the contribution by strong cooperators is on average 18.95 for Turkey, while it is on average 18.44 for Germany ($p = 0.886$).[5]

1.5.3 Punishment behavior

Figure 1.3 shows the mean punishment points for deviations from others' average contribution for Turkey and Germany. The numbers above the bars stand for the relative frequency of observations (in percent) in the respective deviation interval.

[5]There were not any subjects who contribute at least 15 tokens to the public good but do not punish other subjects in the first period for Turkey; therefore, the comparison of contribution levels made by cooperators across Turkey and Germany is not available. Note that the contribution by cooperators is on average 19.17 for Germany.

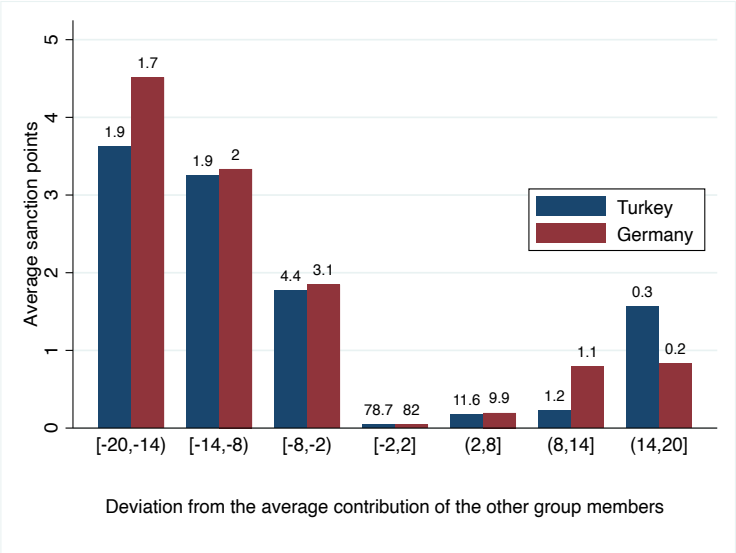


Figure 1.3: Received Punishment Points for Deviations from Others' Average Contribution

We analyze punishment expenditures under two categories: free-rider and anti-social punishment. In the Appendix, Figure A.4 depicts the mean punishment expenditures under these two categories. Free-rider punishment is defined as the sanctioning points assigned to the subjects who contributed less to the public good compared to the punishing person. Anti-social punishment is defined as punishment towards another subject who contributed at least as much as the punishing subject.

Average free-rider punishment is very similar in both countries (1.61 in Turkey, and 1.54 in Germany). Disregarding the cases for punishment expenditure was zero, the free-rider punishment is still similar: on average 2.82 for Turkey and 3.32 for Germany. Free-rider punishment is strictly greater than 0 for 57.24% of all possible cases in Turkey and for 46.46% of possible cases in Germany. To sum up, we do not find any significant differences between the two samples in terms of the free-rider punishment severity and frequency.

The overall anti-social punishment is very similar in both countries, too. It is on average 0.035 for Turkey and 0.037 for Germany. Considering only anti-social punishment that are strictly greater than zero, the average is 1.57 for Turkey, while

it is 1.83 for Germany ($p = 0.037$). The frequency of anti-social punishment does not differ between both countries. Anti-social punishment is executed in 2.24% of all possible cases in Turkey; whereas it is for 2.05% of all possible cases in Germany.

The comparison of anti-social punishment between Turkish samples from endogenous and exogenous settings does not reveal any differences. Anti-social punishment expenditures are on average 0.401 in our Turkish sample and 0.712 in Herrmann et al. [2008] for the first period; whereas they are 0.210 and 0.505 for the first five periods, respectively. The frequency of anti-social punishment does not differ between both samples, either. Anti-social punishment is executed in 26.97% of all possible cases in our Turkish sample and 19.82% of all possible cases in Herrmann et al. [2008] for the first period; while the respective percentages are 14.16% and 18.17% for the first five periods.

In the Appendix, we provide additional information on punishment behavior. Figure A.5 shows the overall average punishment tokens sent per period, in Turkey and Germany, regardless of the contribution deviations. Figures A.6 and A.7 inform about the frequency and the severity of anti-social punishment in Turkey and Germany.

Mean punishment expenditures by different types of subjects

In this section, we compare free-rider and anti-social punishment expenditures by two types of players; free-riders and strong cooperators, as defined in section 1.5.2, across two countries. Accordingly, Table 1.5 denotes the mean level of free-rider and anti-social punishment made by free-riders and strong cooperators, respectively. We do not observe any significant differences of punishment by types, i.e., conveyed by the free-riders or strong cooperators across Turkey and Germany.

In the Appendix, we provide additional information on anti-social punishment behavior of different types of subjects. Figure A.8 and A.10 inform about the frequency and the severity of anti-social punishment conveyed by free-riders. Figure A.9 and A.11 demonstrate the respective statistics for strong cooperators defined according to Table 1.5.

We also categorize subjects into different types according to the level of contribution and punishment in the first three period in Table A.3. In the Appendix, Figure A.12 and A.14 depict the frequency and the severity of anti-social punishment conveyed by free-riders. Figure A.13 and A.15 show the respective statistics for strong cooperators defined according to Table A.3.

Table 1.5: Mean punishment expenditures by different types of subjects

	Free-rider				Strong cooperator			
	Free-rider punishment		Anti-social punishment		Free-rider punishment		Anti-social punishment	
Statistic	TUR	GER	TUR	GER	TUR	GER	TUR	GER
Mean	0.96	1.41	0.081	0.169	2.35	2.78	0.037	0.007
Mean*	1.95	3.31	1.64	1.91	3.43	3.78	1.51	11.00
Frequency (%)	49.55	42.62	4.94	8.84	68.60	73.65	2.47	0.07

TUR stands for the subjects in the respective category (e.g. free riders in Turkey)

GER stands for the subjects in such category (e.g. strong cooperators in Germany)

The statistic mean denotes the average value of the respective punishment

The statistic mean* stands for the average value where the cases of 0 are excluded in the calculations

The statistic frequency denotes the cases where punishment is strictly greater than 0

Result 3 *In Turkey and Germany, free-riders are punished similarly in terms of punishment frequency and severity. The frequency of anti-social punishment is similar in Turkey and Germany, while the severity of anti-social punishment executions is slightly but significantly higher in Germany. In addition, there are no between country-differences with regard to punishment behavior of both free-riders and strong cooperators.*

1.5.4 Population allocation across institutions

In this section, we first compare the popularity of institutions across two countries. The data do not show significant differences on the aggregate level: in the VCM community, the number of subjects is on average 2.95 (24%) for Turkey, whereas it is 3.41 (28%) for Germany. Additionally, in the PUN community, the number of subjects is on average 9.19 (76%) for Turkey, whereas in the PUN community in Germany, the number of subjects is on average 8.59 (72%) for Germany. There are

also no significant differences between both countries if we analyze the popularity of institutions according to different subject types.[6]

Individual change/migration behavior

A closer examination of the switches from PUN to VCM, and vice versa, reveals that German subjects on average switch their communities more often than the Turkish subjects do (Turkey: 2.64, Germany: 4.21, $p = 0.011$). It should be noted that the experiment lasts for 30 periods in Germany whereas the same game lasts for 20 periods in Turkey. If we consider only the first 20 period for Germany, the number of switches becomes 3.61 on average for this sample, still indicating a significant difference in between two samples ($p = 0.032$). The average number of switches per free-rider is 3.38 on average in Turkey, while, in Germany, this value is 9.17 ($p = 0.092$). Also, the average number of switches per strong cooperator is 1.71 on average in Turkey, while, in Germany, this value is 2.57.[7]

In the Appendix, we provide additional information on the dynamics of change behavior. Figure A.16 shows the relative portions of changing and remaining subjects in each community, in each period. One can easily see that the dynamics in Turkey and Germany are very similar.

Community change behavior and next period contributions

We examine the relationship between the received sanction points and the switching behavior from PUN to VCM, and the contribution adjustments when subjects stay

[6]Since we do not have any subjects categorized as cooperators in Turkey, we provide the respective analysis for free-riders and strong cooperators only. In the VCM community, the number of free-riders is on average 0.72 (20%) for Turkey, whereas it is 0.78 (39%) for Germany. Additionally, in the PUN community, the number of free-riders is on average 2.89 (80%) for Turkey, whereas in the PUN community in Germany, the number of free-riders is on average 1.22 (61%) for Germany.In the VCM community, the number of strong cooperators is on average 0.15 (5%) for Turkey, whereas it is 0.21 (8.6%) for Germany. In the PUN community, the number of strong cooperators is on average 3.28 (95%) for Turkey, whereas in the PUN community in Germany, it is on average 2.22 (91.4%) for Germany.

[7]If we consider only the first 20 period for Germany, the number of switches per free-rider becomes 7.50 on average for this sample, still indicating a significant difference between two samples ($p = 0.092$), and the number of switches per strong cooperator becomes 2.29 on average indicating no significant difference.

in PUN. First two columns (models) of Table 1.6 show probit models to identify the one-to-one relation between received punishment and the movement from PUN to VCM community for Turkey and Germany. In the third and the fourth columns, we use GLS random effects model to reflect the determinants of contribution adjustments for the subsequent periods when subjects remain in PUN. The fifth and sixth columns are GLS random effects model that only consider the contributions for periods that subjects switch from PUN to the VCM community.

The dependent variable `Switched from PUN to VCM` is a binary variable that takes the value of 1 if a subject switches from PUN community to VCM community for the next period. The independent variable affecting the community movement is the received sanction points. The first two columns of Table 1.6 show the marginal effects for the percentage point changes when the respective independent variable increases by one unit. Accordingly, the first column of Table 1.6 is the switch from the punishment community to VCM community in Turkey while the second column stands for the switch from PUN to VCM in Germany.

Table 1.6: Community change behavior and next period contributions

	(1)	(2)	(3)	(4)	(5)	(6)
	Switched from PUN to VCM		Next cont when staying in PUN		Cont after switch from PUN to VCM	
	TUR	GER	TUR	GER	TUR	GER
Sanction Points	0.099***	0.097***	0.264***	0.037**	-0.067	-0.099
	(0.013)	(0.010)	(0.029)	(0.018)	(0.115)	(0.081)
Contribution			0.693***	0.426***	0.012	-0.131
			(0.030)	(0.022)	(0.134)	(0.116)
Constant	-1.832***	-1.767***	5.849***	11.182***	5.295**	7.283***
	(0.066)	(0.071)	(0.592)	(0.451)	(2.367)	(2.227)
N	1615	2784	1114	1801	92	173

(1) and (2): Probit Regressions, (3) – (6): GLS random effects models, standard errors in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

The received sanction points increase the likelihood of switching from the PUN community to the VCM community in both countries. Specifically, one point increase in the received sanction points in the punishment community leads to 9.9

percent increase in the likelihood of switching from the PUN community to the VCM community in Turkey while it leads to 9.7 percent rise in the probability of switching in Germany. We conclude that the received sanction points has a similar effect on the switching behavior from PUN to VCM community in both countries.

In the third and fourth columns, we analyze the contribution adjustments for the subjects who were in the PUN community at time $t-1$ and remain in PUN at time t. We use subject's own contribution in addition to the independent variable in the first two columns to explain the adjustments to the contribution for the next period. Both independent variables have a significant impact on the next round's contributions for the subjects that remain in the punishment community. Accordingly, an increase on any of these two variables leads to an increase in the next period's contribution for the subjects that choose to stay in the PUN community (0.264 and 0.037 points in Turkey and Germany, respectively, $p = 0.121$). Apparently, subjects, who do not leave the PUN community, increase their contributions to avoid future sanctions. Furthermore, for subjects staying in the PUN community, one point increase in the subjects' own contribution results in an average increase of 0.693 in the next period's contribution in Turkey, and 0.426 in Germany ($p = 0.366$). For subjects who switch to VCM community, we observe that neither contribution level nor received sanction points are significant to affect the next period contributions (see fifth and sixth columns).

To sum up, sanctions have a similar impact on subjects to leave PUN for VCM in Turkey and Germany. For the switching subjects to VCM, sanctions do not affect their next period contributions, in neither country. For subjects staying in PUN after being punished, sanctions have a significant effect on their next round contributions, with subjects in Turkey showing somewhat stronger reaction than Germans.

Community choice and norms of civic cooperation

To investigate whether there is a relationship between the community choices and the norms of civic cooperation, we ask subjects in Turkey three different questions

to measure their score on civic norms.[8] We set score of 1 as the lowest level of cooperation and the score of 10 as the highest level of cooperation (free-rider) on each question. Then, we sum the scores on these three questions to obtain a total score of the norms of civic cooperation. We observe that the average score of civic norms do weakly significantly differ between subjects who choose the VCM community and those who join the PUN community in the first period (25.57 vs. 24.21, $p = 0.052$).

Table 1.7: Civic norms

	(1)	(2)	(3)
	Contribution	Contribution VCM	Contribution PUN
Total Civic	0.183	-0.027	0.325*
	(0.119)	(0.106)	(0.167)
Constant	2.040***	3.863**	2.987***
	(2.701)	(2.349)	(4.102)
N	72	42	30
R^2	0.025	0.001	0.077

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

As Table 1.7 shows, there is a positive relationship between the first round contributions and the civic norm score in the PUN community. This means subjects in PUN who value civic norms higher contribute more. However, we do not observe this relationship in the VCM community.

[8]In Germany, this questionnaire was not part of the experiment, so we can analyze only the Turkish subjects' behavior. Unfortunately, in one session the questionnaire data are missing, this is why we have in total 72 observations.

1.6 Discussion

Summing up our main findings, in the initial phase, we observe a similar reluctance in both cultures to join the sanctioning institution, with roughly one in four individuals choosing it. The initial contributions, however, differ between Turkey and Germany. Initially, subjects in Turkey contribute lower amounts in both institutions. This observation is in line with the findings of Herrmann et al. [2008], in the case of the exogenous sanction-free (VCM) and sanctioning (PUN) institutions.

Regarding payoffs, participants in Turkey and Germany obtain very similar earnings, both in the VCM and the PUN institution. In the "long-run", the sanctioning institutions are the clear winner of the institutional competition in both cultures, both in terms of contributions and payoffs. This result extends the findings of Gächter et al. [2008] who show the long-run advantage of exogenous sanctioning institutions over the VCM.

As Herrmann et al. [2008] show, antisocial punishment might substantially differ across cultures and this, in turn, can be a major determinant of the differences in the success of the punishment institution. In our study, German and Turkish subject pools behave similarly in terms of punishment behavior. In both cultures, people discipline free-riders similarly, in terms of frequency of punishment, and its severity. One notable exception is that the severity of anti-social punishment is higher in Germany. Due to similarity of initial punishment preferences in both countries to a great extent, we are not able to test whether endogenous choice of institutions have a tendency to equalize antisocial punishment frequencies in populations with different initial preferences for punishment. Additionally, we do not find any significant differences in antisocial punishment behavior between the Turkish cohort of the study by Herrmann et al. [2008] and our Turkish sample, though the former group invest somewhat more in antisocial punishment. We conclude that more research is needed to better understand the role of institution choice in moderating antisocial punishment. In particular, studies involving subject pools that exhibit substantial heterogeneity in terms of initial antisocial punishment frequencies have the potential

to yield novel results regarding the equalizing effect of endogenous institutions.

We do not reject the claims of Herrmann et al. [2008] pointing out the lower success of the exogenously imposed sanctioning mechanism for subjects from Turkey and some other countries. Indeed, Gurdal et al. (2020) also find a similar result. We believe that, in the current experiment, the opportunity to observe the superior outcomes of the sanctioning mechanism over the coexisting VCM mechanism is the main driver of convergence here for the Turkish subject pool.

Limitations

Researchers using experiments for conducting cross-cultural comparisons face certain difficulties, due to inherent differences in personal and institutional characteristics. Thöni [2019] points out that there is no random assignment of participants to treatments in cross-cultural experiments which leads to methodological challenges. He offers various ways to overcome these potential challenges, such as using student pools with minimal international students, employing the same mode of interaction, keeping the financial incentives comparable, etc. We have tried to address these challenges by using the same experimental design and procedures to minimize the potential confounds and framing experimental earnings in terms of tokens for the experiments in Germany and Turkey. In addition, in both of the countries, the overwhelming majority of student participants were local students. Particularly, in the German subject pool, there were virtually no individuals with Turkish descent.

It could be useful to understand how subjects behave in our setting with exogenous assignment to institutions. We have this data for German subjects as reported in Gürerk et al. [2014], but not for the Turkish cohort. In Germany, the results of the exogenous setting are qualitatively the same, i.e., contributions in the sanctioning institution increase while they decrease in the non-sanctioning institution. There are, however, differences in the dynamics. Details of the analysis can be found in Gürerk et al. [2014].

1.7 Concluding remarks

We believe that experimental methodology has a remarkable potential to contribute to the cross-cultural comparison of the success and the prevalence of different institutions. This is mainly due to the ability to exert control over potential confounds [Thöni, 2019] and the ability to fine-tune the degree of cultural variation via access to various combinations of subject pools. Previous studies mainly employ a design where the same institutional arrangement is tested across different subjects pools, and the results are taken as evidence of the effect culture on social behaviors such as altruism, cooperation, and truth-telling. We think that a more accurate picture regarding the interaction of culture and institutions can be obtained by employing experimental designs where (i) knowledge about the past performance of institutions, (ii) endogenous choice between institutions, or both of these are available so that subjects engage in a better-informed decision-making process. In the experiment reported here, endogenous choice of institutions is the critical factor that allows the eventual convergence of subjects from two distinct cultures.

Placing our results in a broader context, with this study, we demonstrate that previously observed uncooperative behaviors in some cultures can be overcome if the right institutions can be chosen by the members of the society. This is in the spirit of the argument made by Acemoglu and Robinson [2013] in the context of the prosperity of nations and institutions: it is not geography or culture which is decisive for the prosperity of a nation but the availability and the quality of its institutions. This line of argument is also backed by recently published empirical work on institutional change in medieval Germany [Dittmar and Meisenzahl, 2020], and experimental evidence from the lab on voting and institutions [Cobo-Reyes et al., 2019].

Chapter 2

THE SEEDS OF SUCCESS: THE PIVOTAL ROLE OF FIRST ROUND COOPERATION IN PUBLIC GOODS GAMES

2.1 Introduction

Achieving cooperation in group settings where individual and collective interests conflict poses a longstanding challenge. This is because individuals are incentivized to act in their own self-interest rather than cooperating for the collective good, which can lead to detrimental outcomes like the tragedy of the commons. Effective cooperation requires aligning individual and group interests through mechanisms like incentives, penalties, communication, and trust-building. But this remains difficult, as individuals often find ways to free-ride if acting selfishly benefits them more than cooperating. Examples include overfishing, public goods provision, and reducing carbon emissions. Public goods experiments utilizing the voluntary contribution mechanism are a classic way to study cooperation in these social dilemma situations. In their most classical form, these experiments involve a group of subjects being matched in small groups and contributing towards a non-excludable group account over multiple periods. When the group size is n, and one unit contribution to the group account increases each member's earnings by p, with $p < 1$, and $np > 1$, the Nash equilibrium is zero contribution, whereas the efficient outcome is full contribution. In these games, individuals generally start with non-zero contributions and converge to equilibrium zero contributions absent incentives [Isaac and Walker, 1988].

In their seminal studies, Fehr and Gächter [2000, 2002] implement a design where subjects play a game of known length, observe each others' contributions each pe-

riod and decide how much to deduct from the earnings of other group members, i.e., *sanction* each other.[1] This deduction is costly for the subject; therefore, it is classified as a form of strong reciprocity (*costly altruism*). When subjects are allowed to assign punishment points to their group members, this can facilitate cooperation and keep the contribution levels high in the experiment. Herrmann et al. [2008] extended this approach by conducting a study comparing the effectiveness of punishment in sustaining cooperation across 16 cities worldwide. They found punishment successfully raises cooperation in Western cities like Boston, Copenhagen, St. Gallen, Zurich, and Nottingham.[2] However, it failed to do so in Istanbul and other similar cities. Istanbul exhibited more antisocial punishment, where high contributors were targeted, and overall had low punishment effectiveness.[3,4]

Our study revisits cooperation and punishment in Istanbul using the same subject pool but a between-subject design. In contrast to Herrmann et al. [2008]'s within-subject design where all subjects play first without punishment, then with punishment, our subjects participate in either the game without punishment (N-experiment) or with punishment (P-Experiment), not both. Under this between-subject approach, we find substantially higher contributions in Istanbul under punishment versus no punishment. However, average contributions still remain below Western city levels.

We identify two critical factors behind cooperation outcomes in the P-Experiment:

[1]Earlier examples where a sanctioning mechanism is used in similar games involve Yamagishi [1986] where subjects could punish the least cooperative group member and Ostrom et al. [1992] where there is a fixed cost of sanctioning and being sanctioned. Since the game length is unknown by the subjects in the latter study, it is not possible to rule out the incentives for reputation behind contribution and sanctioning decisions.

[2]The average contribution rate is 75% or above in all these cities when punishment is available.

[3]Antisocial punishment occurs when a subject retaliates against a group member who contributed more than her to the group account. Herrmann et al. [2008] argue that the extent of antisocial punishment is negatively correlated with contribution levels, and subject pools where antisocial punishment is rare successfully manage to reach high contribution rates.

[4]Other studies also find location-specific differences, like Gächter and Herrmann [2009] showing higher cooperation and prosocial punishment in Switzerland versus antisocial punishment in Russia. Another example is by Buchan et al. [2011], who find cooperation with a global group correlates positively with country globalization levels. For further work on location-specific differences, see Cárdenas et al. [2012] and Lamba and Mace [2011].

1) The level and distribution of first-period contributions exhibit significant heterogeneity across groups. Groups starting at high initial cooperation sustain those levels, while groups starting at low cooperation remain low throughout.

2) Subjects follow simple contribution rules based on prior contributions and sanctions. We estimate these linear rules from the data.

The interaction of these two factors generates strong persistence in contributions over time. To demonstrate this, we conduct a counterfactual experiment using an agent-based model. Feeding the estimated decision rules and actual first-period data accurately reproduces the evolution of contributions over the course of the game.

Analyzing data from Herrmann et al. [2008], we find similar first-period heterogeneity and contribution persistence over time in cities resembling Istanbul. In particular, these stylized patterns are common across cities that share socio-economic similarities with Istanbul, particularly in Athens (Greece), Dnipropetrovsk (Ukraine), Samara (Russia), Minsk (Belarus), Riyadh (Saudi Arabia), and Muscat (Oman). Our Istanbul results likely extend to these cities under a between-subject design.

Our findings suggest that accounting for heterogeneous initial cooperation, rather than just city-level averages, is critical for valid cross-society comparisons. The decay induced by a prior N-experiment in a within-subject design masks actual cooperation capacity. More broadly, populations exhibiting diversity in initial cooperation may see larger gains from institutions designed to enhance cooperation. This study isolates the effect of experimental design and highlights the pivotal role of initial conditions for cooperation outcomes. It provides guidance for designing future cross-society cooperation experiments to produce robust results.

The rest of the paper is organized as follows: in section 2.2, we discuss the details of the experiment. In section 2.3, we report our empirical findings and document prevalent patterns that guide contribution and sanction decision rules. In section 2.4, we propose and report the results of our agent-based model motivated by the estimated decision rules based on prior contributions and sanctions. In section 2.5, we compare our findings in detail with those by Herrmann et al. [2008], and section

2.6 concludes.

2.2 Experimental Design and Procedures

The experiment is based on the design by Fehr and Gächter [2000] and involves two treatments. The N-experiment involves subjects being randomly matched in groups of four and interacting within the same group throughout 10 periods. Each period, subjects are endowed with an initial endowment of 20 tokens, from which they can contribute to a "group project". For every token invested in the group project, each group member earns 0.4 tokens. Subjects' period earnings are calculated as the sum of earnings from the group project and the part of the endowment not invested in the group project. The P-experiment is built on the N-experiment. In addition to the contribution stage, it involves an extra stage where subjects observe the contributions (but not the identities) of all other members in their group and can assign punishment (sanction) points if they wish. Each punishment point costs 1 token to the subject and reduces the target's earnings by 3 tokens. The total reduction in a subject's earnings is limited to the earnings from the contribution stage. We next discuss the details of the two stages.

2.2.1 The Contribution Stage

In the contribution stage, subjects are allocated into groups of four and remain within the same group throughout 10 periods. In each period, subjects are endowed with an initial endowment of 20 tokens, and they simultaneously choose how much to invest in the group project. For each group, the sum of all contributions made to the group project is multiplied by 0.4 and then returned to each group member in the respective group.

If g_i is the individual contribution made by player i and G is the total contribution made to the group project in player i's group, then the total payoff in token terms for player i in the N-experiment is given by:

$$\pi_i = 20 - g_i + 0.4G \tag{2.1}$$

After the contribution stage, subjects are informed about their individual contributions along with the total payoff for all members in their respective groups without disclosing the identity of the subjects. Note that the contribution stage is the only stage of the N-experiment and is repeated for 10 periods in both treatments.

2.2.2 The Punishment Stage

The punishment stage takes place only in the P-experiment. After the contribution stage, all subjects simultaneously make a punishment decision toward other members of their group. For each token player i assigns to player j, the payoff of player i will decrease by 1 token, and the payoff of player j will decrease by 3 tokens.

Let p_{ij} denote the punishment tokens assigned by player i to player j of the same group. If $p_i = \sum_{j \neq i} p_{ij}$ is the total punishment points assigned by player i to other group members and $p_{-i} = \sum_{j \neq i} p_{ji}$ is the total punishment points assigned to player i by other group members, then the total payoff in token terms for player i in the P-experiment is given by:

$$\pi_i = \max\left\{0, (20 - g_i + 0.4G - p_i - 3p_{-i})\right\} \tag{2.2}$$

After the punishment stage, subjects are informed about the individual contributions along with the sent and received punishment tokens and the total payoffs of all members in their respective groups without disclosing their identities. The punishment stage is repeated for 10 periods after the contribution stage during the P-experiment.

2.2.3 Experimental Procedures

The experiment was conducted with 120 subjects in a total of 10 sessions at the Economics Laboratory of Boğaziçi University in Istanbul. Each session involved 12 subjects. We implemented a between-subject design where 60 subjects took part only in the N-experiment, and the other 60 subjects took part only in the P-experiment.[5] An e-mail was sent to subjects who previously showed interest in

[5]Note that this differs from Herrmann et al. [2008], who implement a *within-subject* design for the majority of the subject pools in their sample, including Istanbul. In their design, the

participating in economics experiments. Subjects were able to register online for a date and time of their choosing. No subject participated more than once, and the sessions lasted 45 minutes on average. Subjects were paid in cash at the end of the experiment, and the exchange rate was 0.1 Turkish liras per token.[6]

2.3 Results

We provide summary statistics for the observed contribution levels in Table 2.1. The contributions in our study start at very close levels (around 9) in both the P-experiment and the N-experiment. However, as of the 10^{th} period, the contributions diminish to very low levels (2.85) in the N-experiment while moderately rising to a mean value of approximately 12 in the P-experiment. We find a significant difference between the P-experiment and the N-experiment when all periods are considered ($p = 0.019$) but not for Period 1 ($p = 0.950$). Considering all periods, the computation of statistical power analysis on the respective non-parametric test to reject the equality of the average contributions in the N-experiment and the P-experiment yields a power value of $pw = 0.78$.[7]

Herrmann et al. [2008] employ the design by Fehr and Gächter [2000] to measure the performance of the punishment mechanism among 1120 subjects in 16 different cities worldwide, including Istanbul. We report the contributions of subjects in Istanbul, Boston, and Copenhagen from that study in Table 2.1.[8]

P-experiment follows the N-experiment, whereas, in ours, subjects participate in only one treatment. Other differences between our experiment and theirs are the average session size (21 in Herrmann et al. [2008] and 12 in ours) and the total number of subjects (64 for Istanbul in Herrmann et al. [2008] and 60 in ours).

[6]At the time of the experiments (November 9-11, 2015), 1 Turkish lira corresponded to approximately \$0.35 United States dollars. The interface our subjects encountered was the Turkish translation of the one by Herrmann et al. [2008].

[7]Since the power value of the test is reasonably high at 0.78, it is safe to assume that there is a significant difference in the average contributions between the P-experiment and the N-experiment with our sample sizes. For a detailed discussion on statistical power analysis, see Cohen [2013].

[8]The latter two cities are where the highest average contributions in the P-experiment or the N-experiment were reported, respectively. We would like to iterate that while we implement a *between-subject* design, Herrmann et al. [2008] use a *within-subject* design where the P-experiment follows the N-experiment.

Using the data from Istanbul, Herrmann et al. [2008] find a difference between the two treatments in Period 1 (with the P-experiment generating lower contributions than the N-experiment), but no significant difference when all periods are considered. The summary statistics and non-parametric test results suggest that a between-subject design increases the effectiveness of the punishment mechanism in Istanbul. However, compared to the results from Boston and Copenhagen, our subjects still contribute at lower rates on average, even under punishment. In addition, compared to Boston and Copenhagen, the availability of the punishment mechanism results in much lower average earnings in Istanbul, both for the current study and for Herrmann et al. [2008].[9]

Table 2.1: Mean Contributions and Earnings

	Contribution in Period 1			Contribution in All Periods			Earnings	
	N	P	*p*-value	N	P	*p*-value	N	P
Istanbul - This study	8.9	9.1	0.950	6.1	10.9	0.019	23.6	19.4
Istanbul - Herrmann et al. [2008]	8.9	6.5	0.034	5.4	7.1	0.326	23.3	17.0
Boston - Herrmann et al. [2008]	13.0	16.0	0.012	9.3	18.0	0.002	25.6	27.9
Copenhagen - Herrmann et al. [2008]	14.1	15.4	0.088	11.5	17.7	0.001	26.9	27.7

Notes: The table reports the average contributions (out of 20 tokens) and the p-values for Mann-Whitney tests that use group average contributions as independent observations and test for the equality of contributions in the N-experiment and the P-experiment.

We next report and discuss the results from our two treatments.

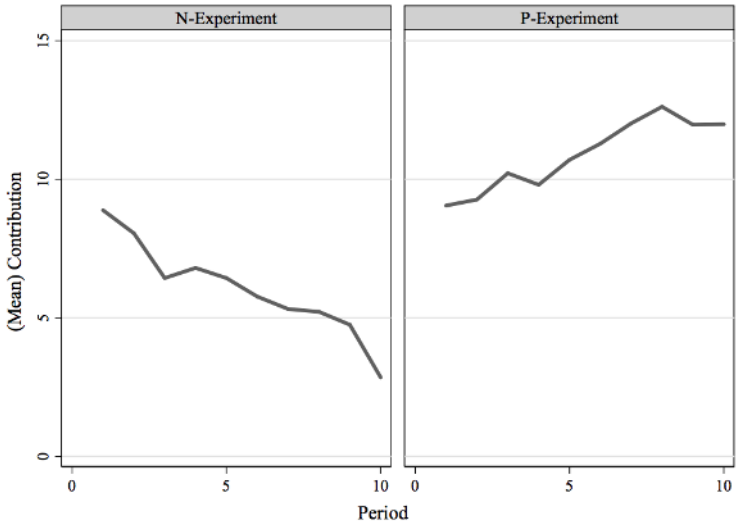
2.3.1 N-Experiment

The overall behavior in the N-experiment exhibits the classical pattern of decaying contributions. We present the evolution of contributions averaged over groups for the N and P-experiments in Figure 2.1. During the first half of the N-experiment, the frequency of subjects who contribute zero is 27%, whereas in the second half, this frequency reaches 46%. Among the 15 distinct groups in this experiment, 11

[9]For a detailed comparison, see section 5.

groups end up at very low (3 tokens or fewer) average contribution levels, 3 groups end up at moderate (between 6.75 and 8.5 tokens) contribution levels, and only one group manages to reach a contribution level of 12 by the 10^{th} period. While these values point out some degree of heterogeneity, clearly the most dominant pattern observed in the N-Experiment is the decline of contributions over periods. On the contrary, the evolution of contributions in the P-Experiment exhibits a great deal of heterogeneity among groups, as discussed in detail in the next section.[10]

Figure 2.1: Timeline of Average Contribution by Treatment



Notes: The left panel reflects the average contribution in N-experiment while the right panel shows the respective statistic in P-experiment.

2.3.2 P-Experiment

We observe some salient patterns from the P-experiment. One decisive pattern is that the way group members update their contributions in the P-experiment follows a close relationship with the respective group average. In particular, when a group

[10]In the first period, the mean (and standard deviation) of average contributions in the N-Experiment was 8.88 (6.66), and in the P-Experiment, it was 9.05 (6.49). By the tenth period, these statistics had changed significantly, with the mean contribution in the N-Experiment decreasing to 2.85 (5.42) and the mean contribution in the P-Experiment increasing to 11.98 (6.77).

Table 2.2: Contribution Change by Group Average

Group Average	Next Period Contribution		
	Increases	Remains the Same	Decreases
Lower	39 (16.5%)	81 (34.3%)	116 (49.2%)
Equal	9 (13.4%)	53 (79.1%)	5 (7.5%)
Higher	161 (67.9%)	45 (19.0%)	31 (13.1%)

Notes: The table reports the frequency of next-period contribution changes by subjects in the P-experiment when they compare their current-period contribution to the group average. Percentages are calculated by row and represent the proportion of subjects in each category.

member observes that her contribution falls short of the group average, she seldom decreases her contribution next period. Instead, she almost always either increases her contribution or keeps it the same. Similarly, when her contribution exceeds the group average, the next period's contribution is frequently either lower or the same. We report the frequencies of the direction of change in contributions conditional on how one's last period contribution compares to the group average in Table 2.2.

Decision to Change Contributions

We run a series of logit and OLS regressions to unveil the driving forces behind the contribution decisions.

In a set of logit regressions reported in Table 2.3, we first set out to understand the dynamics behind decisions to revise (increase or decrease) contributions in the P-experiment. Here, we also control for certain attitudinal and demographic information about subjects and capture group-level fixed effects.[11]

[11]The set of control variables used in regressions involve the following: age in years, a gender dummy, number of older siblings, number of younger siblings, a dummy for being an only child, binary response to the trust question ("Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"), Likert response to the risk question ("How willing are you to take risks in general?"), the number of economic classes taken (censored at 4), number of friends in the session, an indicator for membership in an organization, Likert response to the question "How reliable do you think your responses are

In Model 1 and Model 2, the dependent variable, *change*, takes the value of 1 if a subject's contribution in the current period differs from her contribution in the previous period and 0 otherwise. Our estimations reveal that the only variable significantly affecting the *change* variable is the punishment points received from other group members in the previous period (Model 1). This effect is robust to controlling for other factors (Model 2).

In Model 3 and Model 4, the dependent variable, *increase*, takes the value of 1 if a subject's contribution in the current period is larger than her contribution in the previous period and 0 otherwise. In these models, we restrict our working sample to a subset of observations featuring only subjects contributing less than the previous period's group average. Once again, the only variable with a significant effect is the number of punishment points received from other group members in the previous period (Model 3). This effect is also robust to controlling for other factors (Model 4).

In Model 5 and Model 6, the dependent variable, *decrease*, takes the value of 1 if a subject's contribution in the current period is less than her contribution in the previous period and 0 otherwise. In these models, we again restrict our working sample to a subset of observations, this time featuring only subjects contributing more than or equal to the previous period's group average. In these estimations, the average contribution of other group members in the previous period and contribution in the previous period significantly affects the next period's contributions (Model 5). When controlling for other factors (Model 6), punishment points received from other group members in the previous period also yield a significant effect.

After investigating the conditional frequencies of change decisions, we next scrutinize the intensive margin, i.e., how much contributions vary in magnitude by these factors.

---

in this experiment?"

Magnitude of Change in Contributions

We next run a series of ordinary least squares (OLS) regressions and report our findings in Table 2.4. Here, we use *contribution* as the dependent variable and restrict our working sample to observations featuring subjects who revise their contribution in the current period. Models 1 and 2 involve all subjects who either raised or reduced their contributions (but not those who kept it the same). The number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period, and the last period's contribution have a significant effect on the next period's contributions (Model 1). The effect of these variables preserves significance when controlling for other factors (Model 2).

In Models 3 and 4, we restrict our working sample to observations featuring subjects who contribute less than the group average in the previous period and raise their contributions in the current period. Here, the number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period, the last period contribution, and the period variable has a significant effect on the next period contributions (Model 3). The significance of these variables (except period) is robust to additional controls (Model 4).

In Models 5 and Model 6, we restrict our working sample to observations featuring subjects who contribute more than or equal to the group average in the previous period and reduce their contributions in the current period. The number of punishment points received from other group members in the previous period, the average contribution of other group members in the previous period, and the last period's contribution have a significant effect on the next period's contributions (Model 5). Not surprisingly, received punishment points have a negative effect on contributions since they constitute antisocial punishment in this setting. The effect of these variables preserves significance when controlling for other factors (Model 6).

Table 2.3: Determinants of Contributions Changes (LOGIT)

	(1)	(2)	(3)	(4)	(5)	(6)
Received Points in $t-1$	0.314**	0.249**	0.175*	0.214**	0.111	0.190**
	(0.124)	(0.108)	(0.100)	(0.093)	(0.076)	(0.089)
Other's average contribution in $t-1$	-0.047	-0.047	0.043	0.057	-0.318***	-0.342***
	(0.044)	(0.038)	(0.064)	(0.047)	(0.066)	(0.059)
Contribution in $t-1$	0.017	-0.005	0.035	-0.010	0.216***	0.265***
	(0.029)	(0.037)	(0.077)	(0.060)	(0.040)	(0.046)
Period	0.084	-0.020	-0.068	-0.006	-0.018	-0.039
	(0.072)	(0.055)	(0.054)	(0.075)	(0.067)	(0.078)
Final period	-0.395	-0.067	-0.034	-0.227	0.376	0.392
	(0.288)	(0.242)	(0.573)	(0.527)	(0.351)	(0.468)
Constant	No	No	No	No	No	No
Controls	No	Yes	No	Yes	No	Yes
N	540	540	237	237	303	303
Adj. R^2						

Notes: Standard errors are clustered across different groups, reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: Determinants of Contributions (`OLS`)

	(1)	(2)	(3)	(4)	(5)	(6)
Received Points in $t-1$	0.244**	0.191*	0.399***	0.324**	-0.513**	-0.363**
	(0.089)	(0.108)	(0.107)	(0.119)	(0.187)	(0.153)
Other's average contribution in $t-1$	0.612***	0.608***	0.473***	0.426***	0.495***	0.565***
	(0.067)	(0.067)	(0.090)	(0.107)	(0.159)	(0.162)
Contribution in $t-1$	0.336***	0.256***	0.511***	0.493***	0.412***	0.341**
	(0.073)	(0.073)	(0.097)	(0.099)	(0.118)	(0.125)
Period	0.039	-0.031	0.151*	-0.028	-0.107	-0.111
	(0.061)	(0.125)	(0.082)	(0.124)	(0.069)	(0.207)
Final Period	-0.527	-0.262	-0.649	0.007	0.471	0.198
	(0.603)	(0.768)	(0.528)	(0.633)	(1.228)	(1.601)
Constant	No	No	No	No	No	No
Controls	No	Yes	No	Yes	No	Yes
N	361	361	161	161	121	121
Adj. R^2	0.895	0.901	0.956	0.961	0.880	0.898

Notes: Standard errors are clustered across different groups, reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5: Social versus Antisocial Punishment

Sender's Contribution	Punishment			
	Non-zero	Zero	Mean	Median
Higher	338 (52.3%)	308 (47.7%)	1.09	1
Not Higher	204 (17.7%)	950 (82.3%)	0.33	0

Notes: The table reports the frequency of social and antisocial punishments along with the mean and median statistics.

Punishment Decisions

Table 2.3 and 2.4 demonstrate that punishment matters in contribution decisions. We next investigate the frequency and determinants of punishment decisions.

In Table 2.5, we report punishment frequencies along with mean punishment points conditional on the relative contributions of subjects sending the punishment (sender) and target subjects (receiver). While punishment expenditures are often not high, subjects punish those who contributed less than them around 52% of the time. On the contrary, antisocial punishment is not prevalent and occurs only in around 18% of cases when a subject observes that a group member contributed at least as much as herself.[12]

We next disentangle the effects of different variables on the likelihood of engaging in punishment. Since punishment expenditures are often low and zero-punishment commonly observed, we chose to use a series of probit models instead of linear regressions and report our results in Table 2.6.

Model 1 (Model 3) and Model 2 (Model 4) use the first-period observations in the P-experiment. While the latter model involves a set of attitudinal and demographic variables as controls, the former one does not. When we restrict our working sample to observations in which the receiver's contribution is less than that of the sender, the two strongly significant determinants of social punishment in the first period are

[12]Note that the mean punishment expenditures are very close to the values obtained in Herrmann et al. [2008] for Istanbul.

the *contribution of the target subject* (i.e., receiver) and the *average contribution of the two remaining group members* (Model 1). The significance of these variables is robust to additional controls (Model 2).

Models 3 and 4 consider instances, where the sender's contribution is less than or equal to the receiver in the first period. We see that only the contribution of the target subject (i.e., receiver) significantly affects anti-social punishment in the first period; however, the significance of this variable vanishes when we control for other factors.

Model 5 and Model 6 use all observations in the P-experiment, and again while the latter involves a set of attitudinal and demographic variables as controls, the former does not. We document that the sender's contribution, the average contribution of the two remaining group members, the punishment points that the subjects received in the previous period, the receiver's contribution, and the period variable have an effect on the punishment decision (Model 5). All these variables except the sender's contribution remain significant determinants of punishment decisions controlling for other factors (Model 6).

Using a linear regression model, Herrmann et al. [2008] find that the amount of social punishment is negatively related to the contribution of the target subject and positively related to the average contribution of remaining group members in Istanbul (`Table S3A`). On the other hand, they find that the amount of antisocial punishment for the same subject pool is negatively related to the punisher's contribution and the experiment period (`Table S3B`). For the current study, we obtain the same sign for all the effects, but the effect of the punisher's contribution to antisocial punishment is insignificant. In addition, they do not find an effect of the punishment points received in the previous period on antisocial punishment. In contrast, we find a positive effect of this variable in this study.

We next propose an agent-based model that combines the salient empirical decision patterns of our experiment with first-period period contributions.

Table 2.6: Punishment Decisions

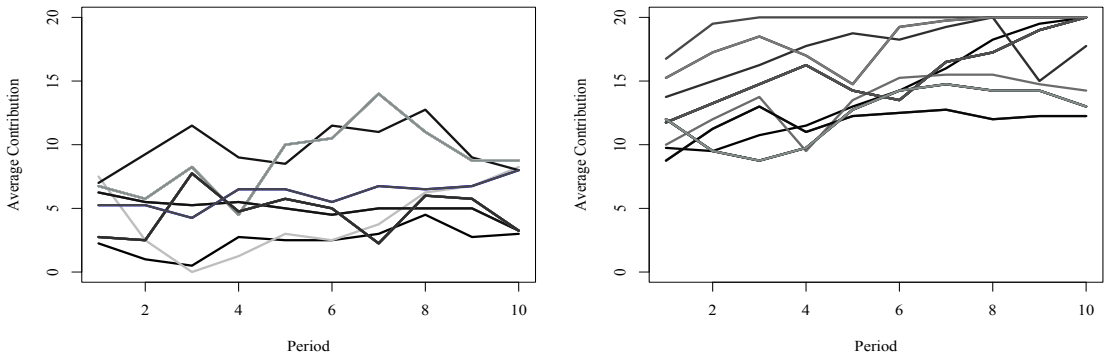
	(1)	(2)	(3)	(4)	(5)	(6)
Receiver's Contribution	-0.100***	-0.151***	-0.085***	-0.018	-0.146***	-0.156***
	(0.034)	(0.051)	(0.030)	(0.035)	(0.021)	(0.023)
Sender's Contribution	0.007	-0.048	0.026	-0.023	0.033*	0.025
	(0.032)	(0.058)	(0.041)	(0.050)	(0.020)	(0.024)
Avg. Contr. of Remaining Members	0.067**	0.102**	-0.015	0.078	0.094***	0.106***
	(0.031)	(0.042)	(0.027)	(0.049)	(0.021)	(0.020)
Received Sanction Points in $t-1$					0.045**	0.045*
					(0.022)	(0.023)
Period					-0.082***	-0.063**
					(0.028)	(0.029)
Final Period					0.184	0.082
					(0.161)	(0.188)
Controls	No	Yes	No	Yes	No	No
N	80	80	100	100	1620	1620

Notes: Standard errors are clustered across different groups, reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4 Agent-Based Modeling

In addition to differences arising from the subject design of the P-experiment, we argue that another critical factor that has first-order implications is the heterogeneity in the first round, which has long-lasting consequences.[13] Our empirical findings reveal that if subjects in Istanbul happen to start the P-experiment in a collaborative group, they sustain collaboration rates as high as in Boston and Copenhagen. We highlight the importance of the first-period results in Figure 2.2, where we cluster groups with respect to their first-period average contributions (lowest 7 on the left and highest 8 on the right panel) and plot the evolution of group averages over periods. Figure 2.2 reveals that if a group's average contribution rate is above 43.75% in the initial round, its average contribution rate by the end of the P-experiment is no less than 61.25%.

Figure 2.2: Evolution of Average Contributions in the P-experiment



Notes: The figure on the left shows the average contribution over the horizon of 10 periods for 7 groups with the lowest first-period average contributions whereas the figure on the right demonstrates the statistic for the remaining 8 groups.

In order to elicit the decisive role of the first-period results, we rely on an agent-

[13]Specifically, we argue that drawing conclusions on the success rate of the P-experiment by concentrating only on *group averages* as in Herrmann et al. [2008] can be misleading since this approach smooths out the decisive role of first-round results by aggregation.

based modeling approach. In doing so, we start by taking the first-period contributions from the P-experiment data. Next, for our subjects, we impose a simple state-dependent set of linear decision rules motivated by the data, and we run Monte Carlo simulations for each of the 15 groups in our sample. Specifically, we proceed as follows:

1. We start our counter-factual analysis by feeding the agent-based model with *actual* first-period contribution data of our 60 subjects in the 15 groups.

2. Subjects either change their contribution or make the same amount of contribution in the next period. Following our empirical findings, we link decision rules to subjects' comparison of their last period contribution to the group average: if a subject's last period contribution is less than the group average, she either raises her contribution or keeps contributing the same amount in the next period; and if her last period contribution is at least equal to the group average, she either reduces her contribution or contributes same. In the former case, we rely on a two-layer contribution rule: in the first layer, using estimated probabilities defined over received sanction points last period $s_{i,t-1}$, we simulate if one retains her contribution or not (Model 3, Table 2.3). If she does not retain, in the second layer, we impose that her contribution follows $c_{i,t} = 0.511 \times c_{i,t-1} + 0.473 \times \overline{c}_{-i,t-1} + 0.399 \times \sum_{j \neq i}^{4} s_{j,i,t-1} + 0.151 \times t$. On the contrary, if a subject's last period contribution is greater than or equal to the group average, we similarly rely on a two-layer contribution rule: in the first layer, using estimated probabilities defined over the average contribution of other members in the group during the previous period $\overline{c}_{-i,t-1}$ and last period contribution $c_{i,t-1}$, we simulate if one retains her contribution or not (Model 5, Table 2.3). If she does not retain, in the second layer, we impose that her contribution follows $c_{i,t} = 0.412 \times c_{i,t-1} + 0.495 \times \overline{c}_{-i,t-1} - 0.513 \times \sum_{j \neq i}^{4} s_{j,i,t-1}$.

3. We observe that subjects punish others either by a *single* sanction point or they do not punish others in 85.06% of all possible cases. We see that subject

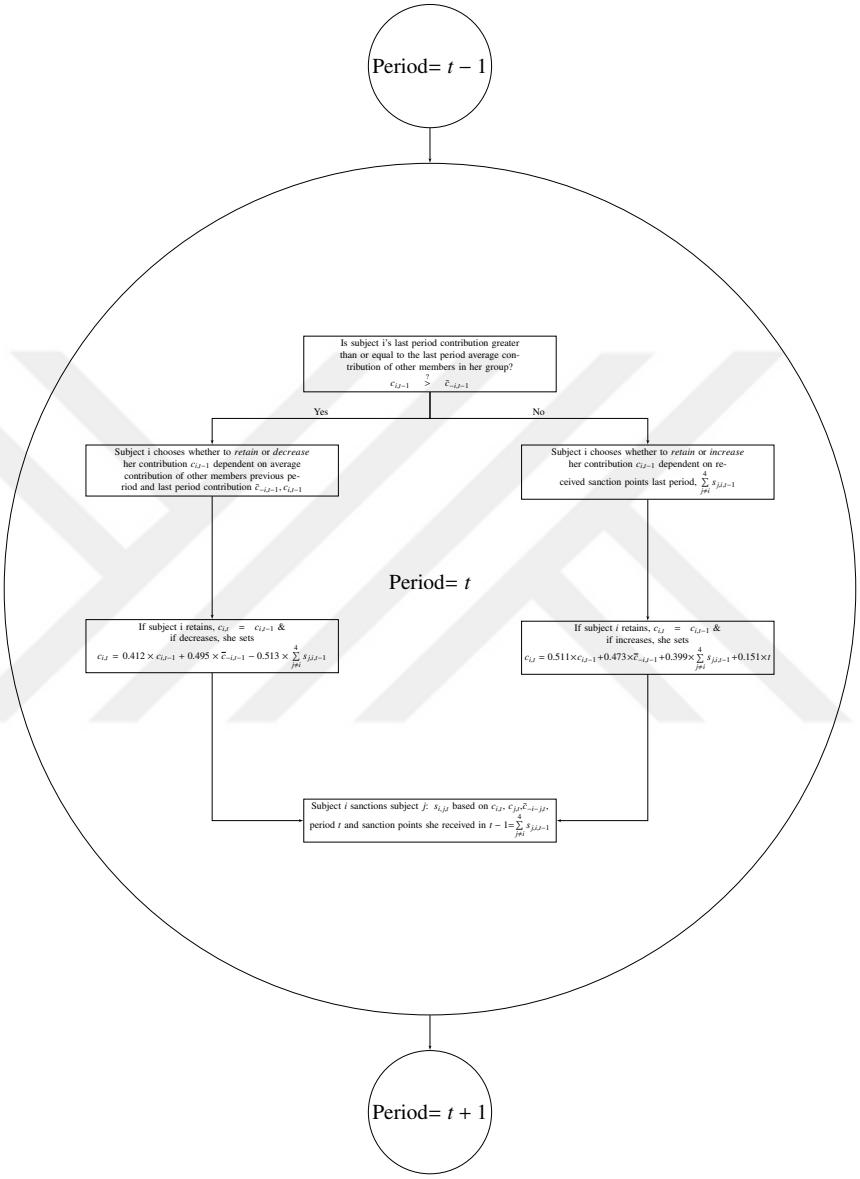
i's sanctioning decision on subject j, $s_{i,j,t}$ depends significantly on subject i's received sanction points last period $s_{i,t-1}$, the average contribution of the remaining members in the respective group in the current period $\overline{c}_{-i-j,t}$ and the round of the current period t, as well as her and subject j's contribution in the current period: $c_{i,t}$ and $c_{j,t}$. Based on the estimated sanction probabilities from the data, we simulate if subject i assigns subject j a sanction point (of unity) or not (Model 5, Table 2.6).[14]

4. We move on to the next period and recursively conduct the same steps over periods.

We summarize the timeline of the agent-based modeling algorithm in Figure 2.3. We display average group contributions by our agent-based model simulations in Figure 2.4. In doing so, for each of the 15 groups, we plot group-specific Monte-Carlo simulation averages (dotted lines) along with their 2-standard-deviation confidence intervals (shaded grey areas), and we display them jointly with *actual* group average contributions (straight lines). Figure 2.4 reveals that despite its simplicity, the decision rule we impose on subjects not only mimics final-round average contributions accurately but also does a fairly good job of capturing the evolution of group averages in most of the instances.

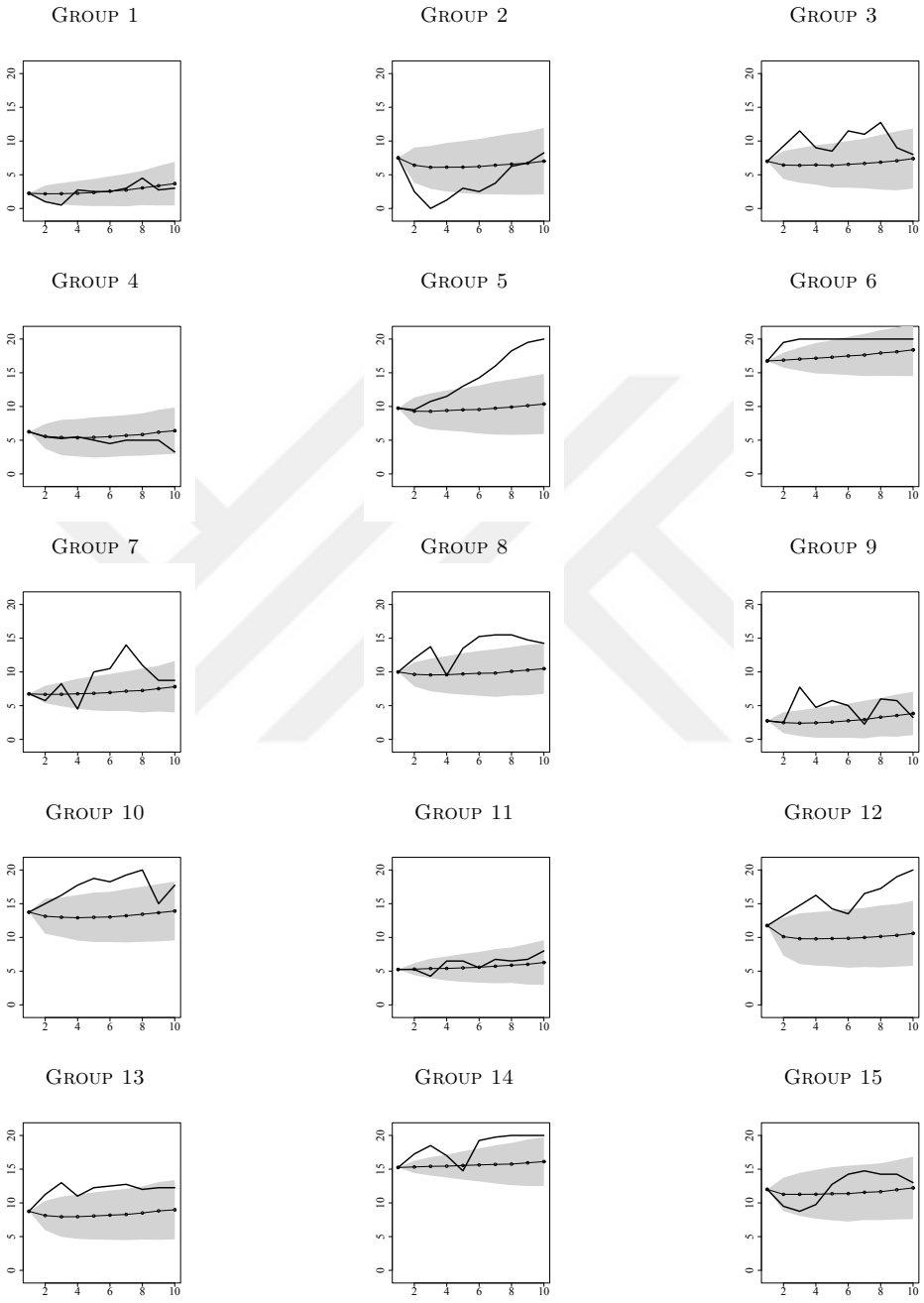
[14]We do not have a history of received sanction points in the first period, so we proceed differently to determine the first-period sanction points. In the first period, we see that subject i's sanctioning decision on subject j, $s_{i,j,1}$ significantly depends on the average contribution of the remaining members in the respective group $\overline{c}_{-i-j,1}$ and j's contribution $c_{j,1}$ for social punishment case (Model 1, Table 2.6), whereas this decision only depends on j's contribution $c_{j,1}$ for anti-social punishment case (Model 3, Table 2.6). Based on estimated sanction probabilities from the data, we simulate if subject i assigns subject j a sanction point (of unity) or not.

Figure 2.3: Timeline of the Model



Notes Figure 2.3 displays the timeline of agent-based modeling for the P-experiment. Only first-period contributions are retrieved from the experiment, while the remaining data is constructed through the modeling. Subjects' decisions to retain or change their contribution level in subsequent periods are determined via Table 2.3. In the case of change, the next period's contributions are constructed by benefiting from Table 2.4. Sanctioning decisions are made via Table 2.6.

Figure 2.4: Agent-Based Model Simulation Results



Notes: The horizontal axis denotes periods, and the vertical axis denotes average group contribution. The solid line refers to group averages by the P-experiment data, the dotted line refers to group-specific average agent-based model results by Monte-Carlo simulations, with the shaded gray areas referring to resultant 2-standard-deviation confidence intervals.

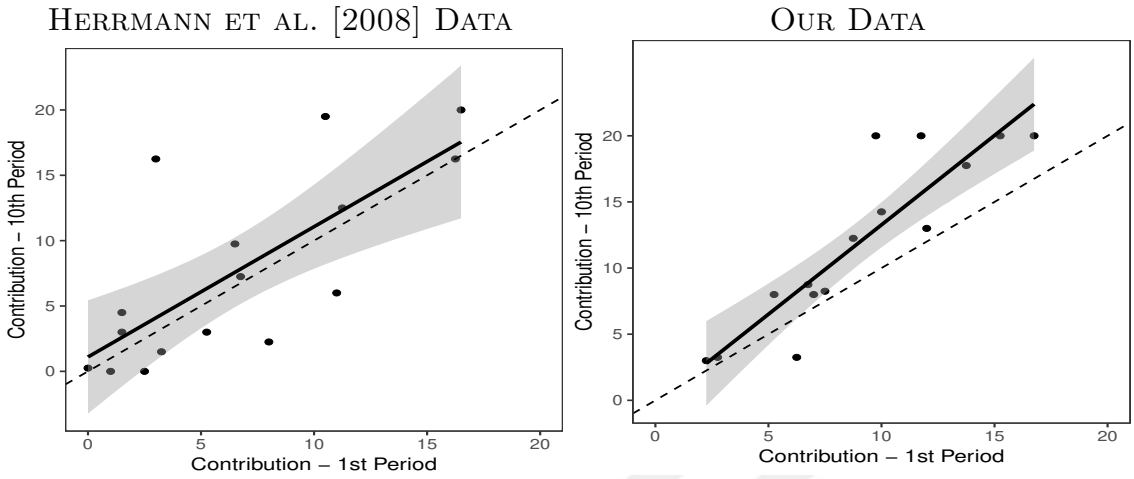
2.5 Comparison with Herrmann et al. [2008]

Heterogeneity in first-period average contributions and its persistence throughout the game is not peculiar to our experiment. Consequently, eliminating the detrimental effects of the N-experiment due to the within-subject design could extend beyond Istanbul.

To demonstrate how our findings compare to those by Herrmann et al. [2008] for Istanbul, we plot first-period average group contributions against their last-period counterparts by our and Herrmann et al. [2008]'s data in Figure 2.5. This depiction reveals that first-period average group contributions by Herrmann et al. [2008] also exhibit a high degree of variability, as do our data. There are, however, two stark differences: First, while only 2 (out of 15) groups in our experiment contribute at a rate lower than 25% in the first period, 7 (out of 16) groups in Herrmann et al. [2008]'s experiment start the game at a contribution rate below 25%. Second, while the slope between the first and last period average group contributions via Herrmann et al. [2008]'s data is close to unity (0.997, with a standard error of 0.243), the slope via our experiment is sizably steeper (1.352, with a standard error of 0.186). This suggests that while average group contributions by Herrmann et al. [2008] stagnated on average over time, groups in our experiment managed on average to raise their average group contributions over periods. Given that our experiment was conducted at the same university to the same subject pool, we argue that these disparities stem from the difference due to our *between-subject* design versus Herrmann et al. [2008]'s *within-subject*, by the last period of the no-punishment treatment of which all average group contributions converged to zero.

Istanbul was not the only city that exhibited sizable heterogeneity in average first-period contributions that persisted for the rest of the P-experiment. The experiments by Herrmann et al. [2008] were conducted in various cities with stark socio-economic differences. To explore which of these cities resemble Istanbul better, we rely on a principal component analysis (PCA) using (i) straight-line physical distance to Istanbul, (ii) GDP per capita (in 2017 current US Dollars), and (iii)

Figure 2.5: Comparison of Average Contributions in the P-experiment



Notes: The unit of observation is the average group contribution. The horizontal axis denotes the 1st-period mean contribution, and the vertical axis denotes the 10th-period average contribution in P-experiment. The points below the dashed 45-degree line imply that those groups failed to improve upon the 1st period mean contribution, while the points above the line show the groups that improved their mean contribution in the final period. The solid line denotes the linear fit, while the shaded gray areas display 95% confidence intervals.

cultural and psychological distance to Istanbul (Turkey) via Muthukrishna et al. [2018]'s `WEIRD` scale index. Our PCA exercise reveals that a cluster of six cities by Herrmann et al. [2008] resembles Istanbul the most: Athens (Greece), Dnipropetrovsk (Ukraine), Samara (Russia), Minsk (Belarus), Riyadh (Saudi Arabia), and Muscat (Oman).[15]

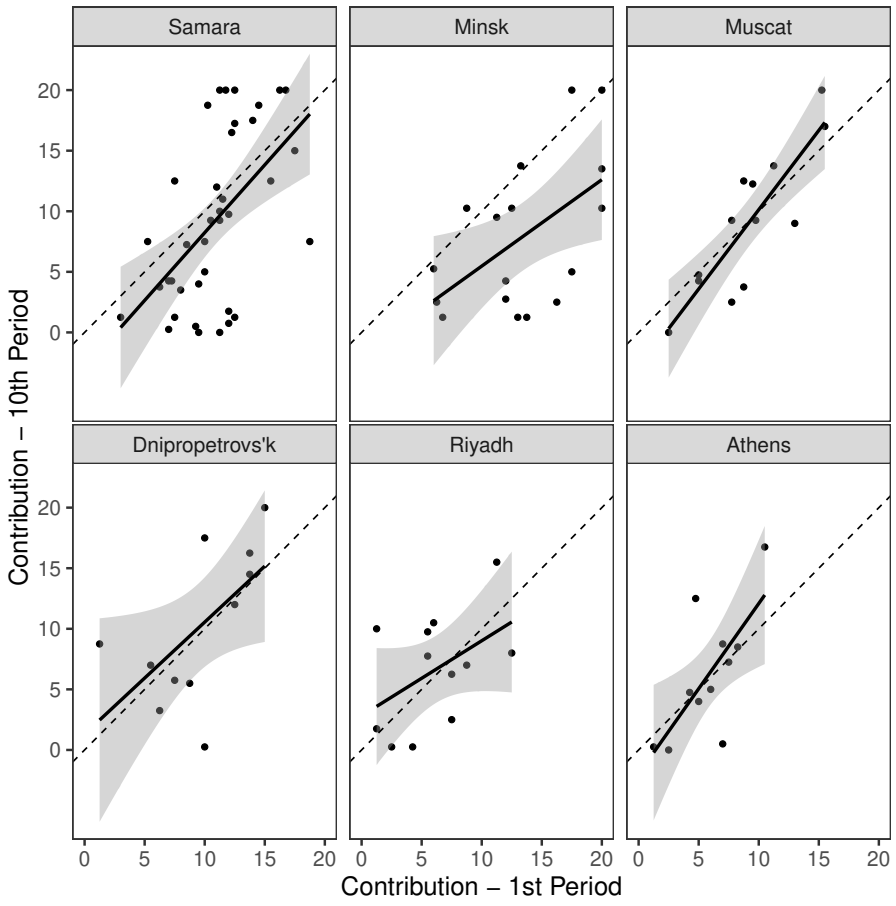
We depict the first-period average group contributions against their last-period counterparts via Herrmann et al. [2008]'s data for these six cities in Figure 2.6. We observe that in all of these cities, there is sizable first and last-period heterogeneity in average group contributions, along with a strong positive correlation between the two, which is close to unity. We further document that a (city) fixed-effect

[15]Four of the cities by Herrmann et al. [2008] lack Muthukrishna et al. [2018]'s `WEIRD` scale index scores. For the lacking cities, we imputed as follows: for Athens of Greece, we proxied via Cyprus; for Riyadh of Saudi Arabia and Muscat of Oman, we proxied via neighboring Gulf countries; and for Copenhagen of Denmark, we proxied via Sweden and Norway. The resultant Figure B.2 is in the Appendix. Details of our PCA exercise are available upon request.

regression of last-period average group contributions on first-period average group contributions by these six cities yields a slope coefficient of 0.975 (with standard error 0.140), which is not statistically different than the slope of 0.997 (with standard error 0.243) that we estimate for Istanbul.[16] As a result, we conclude that Istanbul does not single out with its idiosyncrasy vis-à-vis heterogeneity in first-period average contributions and their persistence throughout the game. Instead, these stylized patterns are statistically common across cities that share socio-economic similarities with Istanbul. As such, the increase in Istanbul's contribution rates due to the between-subject design could plausibly extend to these cities with similar characteristics.

[16]The city fixed-effect regression of the six cities yields a constant coefficient of -1.139 with a standard error of 1.467 whereas the constant coefficient is 1.100 with a standard error of 2.022 for Istanbul. The respective Figure B.1 is in the Appendix.

Figure 2.6: First-Period and Last-Period Contributions by City [Herrmann et al., 2008]



Notes: The unit of observation is the average group contribution. The horizontal axis denotes the 1st-period mean contribution, and the vertical axis denotes the 10th-period average contribution. The points below the dashed 45-degree line imply that those groups failed to improve upon the 1st period mean contribution, while the points above the line show the groups that improved their mean contribution in the final period. The solid line denotes the linear fit, while the shaded gray areas display 95% confidence intervals.

2.6 Discussion and Concluding Remarks

In this paper, we demonstrate that in a public good game with punishment, contributions are significantly higher in Istanbul under a between-subject design compared

to a within-subject design where the no-punishment condition precedes the punishment condition. This highlights the detrimental effect of prior experience without punishment on cooperation. However, average contributions remain below Western city levels, indicating that limited cooperation persists even with sanctions.

We identify two critical factors behind cooperation patterns: heterogeneous initial contributions, extending Burlando and Guala [2005], and simple contribution updating rules based on prior actions estimated from the data. An agent-based model verifies that the interaction of these factors generates strong persistence in contribution levels over time.[17]

The data analysis from Herrmann et al. [2008] reveals similar heterogeneity and persistence over periods in cities resembling Istanbul. As such, our results likely generalize if employing a between-subject design in these settings as well, eliminating the cooperation decay induced by a prior no-punishment condition.

Our results underline the importance of initial contributions in shaping subsequent cooperation within a group, resonating with studies demonstrating the efficacy of grouping subjects based on contribution levels. For instance, Gunnthorsdottir et al. [2007] find lower contribution decline when matching on prior actions versus random matching. Related studies include Gächter and Thöni [2005], Ones and Putterman [2007], and Gunnthorsdottir et al. [2010]. Brekke et al. [2011] also enable endogenous group formation based on charity donations, finding improved cooperation among donors.

As in Gächter et al. [2010], we posit that culture affects cooperation via beliefs and punishment responses. The variance we observe in initial contributions underscores the power of beliefs, aligning with reduced contributions following a within-subject design as in Herrmann et al. [2008]. Personal risk preferences also influence first-round contributions, which then adjust based on received sanctions. A group starting with low contributions coupled with antisocial punishment risks cooperation failure.

[17]See also Andreoni and Miller [2002], who show that altruistic behavior is rationalizable and consistent but also highly heterogeneous across individuals.

Our study makes several contributions. It demonstrates the pivotal impact of experimental design and initial conditions on cooperation outcomes. It provides guidance for robust cross-society experiment design by underscoring within-group heterogeneity. Our findings uniquely isolate the effect of first-round divergence, complementing research on culture and conditional cooperation in social dilemmas. The insights into contribution updating rules and belief formation advance theoretical understanding of how cooperation evolves.

Future work should further explore the sources of heterogeneous initial contributions and beliefs across individuals. Overall, highlighting within-society variation rather than just cross-society differences is critical for drawing valid inferences and crafting policies to encourage cooperation. This study demonstrates the inadequacy of only considering city-level aggregates when cooperation hinges fundamentally on initial beliefs within subgroups.

Chapter 3

SIGNALING YOURSELF

3.1 Introduction

Social learning theory states that people learn through ways of observing and imitating the behaviors of others [Bandura and Walters, 1977]. In their most classical form, the social learning experiments include the nature drawing randomly between two states with equal probability and a sequence of subjects receiving independent and private information with respect to the state of the world to announce public predictions sequentially [Anderson and Holt, 1997]. While in such settings the object of interest is binary and self-irrelevant to the subjects in the network, the response to a private signal might be different when the predicted item is self-relevant to at least one participant in the network. As an example, the assessment of agents' performance within organizational contexts typically involves a sequence of events: their initial self-evaluation, the subsequent response from principals based on that self-assessment, and finally, the revised evaluation of agents.

The foundational role of initial belief about a quality is pivotal in rational decision-making under uncertainty. Yet, people frequently display overconfidence in their beliefs across various domains. For instance, Svenson [1981] documents that more than two thirds of Swedish and U.S. students find themselves better than the median driver. Barber and Odean [2001] finds that excessive trading of common stock investments decreases the return of both men and women investors. Camerer and Lovallo [1999] attributes the high rate of new business failures to being overconfident.

With respect to the comparison of the beliefs across self-relevant and self-irrelevant contexts; Eil and Rao [2011] as well as Möbius et al. [2014] find that individuals pro-

cess information asymmetrically by giving more weight to good news compared to the bad news when it has personal relevance. Ertac [2011] also finds asymmetric processing in the self-relevant context, but asserts that asymmetric processing involves placing more weight to the bad news, resulting in overly pessimistic beliefs. On the other hand, Coutts [2019] does not observe any differences in belief updating between contexts with valence and those without.

When it comes to the comparison of the beliefs about one's own and others' performance, Grossman and Owens [2012] reveals that participants accurately assess others' scores but tend to overestimate their own scores. On the other hand, Bordalo et al. [2019] posits that participants do not only overestimate their own ability, they also overestimate the ability of others.

This paper experimentally investigates several fronts. First, we analyze whether the network structure has any effect on the results when the predicted item is self-irrelevant to the participants in the network. We do so by comparing the deviation of the predictions of the last-mover from the Bayesian benchmark prediction across the different network structures. [1][2] Second, we explore the influence of the prior beliefs on the subsequent beliefs in a social network within a self-relevant context. Specifically, we examine whether there is a difference between the initial beliefs and the subsequent beliefs upon receiving an information. Third, we analyze the response to relative performance feedback across self-relevant and self-irrelevant contexts within the framework of a social network. In particular, we aim to identify whether individuals exhibit overconfidence in terms of overestimating their relative performance, whether there is a difference in the beliefs across self-relevant and self-irrelevant contexts and lastly whether these differences persist.

To measure how individuals make their predictions when the predicted item is not self-relevant to any agent in the network, we first ask subjects to predict a randomly

[1]While Grimm and Mengel [2020] vary the network structure and the information about the network across treatments and show the significance of the information about the network structure, we only vary the network structure so that subjects have perfect information about the network structure before the beginning of all treatments in our experiment.

[2]There is an extensive body of literature [Centola, 2010; Kearns et al., 2006; Watts and Strogatz, 1998] emphasizing the effect of the network structure on the dissemination of information.

chosen target number that ranges between 0 and 100 in a social network. We use two different setups: linear and circular network. In a linear network, three players receive a private signal about the target number and make a prediction sequentially. The private signal will be the comparison of the target number to a randomly drawn number between 0 and 100 as opposed to a binary signal. Every player other than the first-mover will also receive an information about the prediction of the previous players. On the other hand, in a circular network, two players receive a private signal about the target number and make a prediction consecutively. Following that, we provide the first mover an information about the prediction of the second mover and ask him to revise his prediction.

Assuming that subjects are rational, risk-neutral agents and have rational expectations about other subjects, we observe that there is not any significant difference between the subjects' prediction and the corresponding Bayesian benchmark prediction[3]. A closer look at the predictions, however, reveals that subjects underestimate the corresponding Bayesian benchmark prediction with high signal and overestimate it with the low signal when they are lined up in a linear fashion[4]. A possible explanation for such behavior is provided by Edwards [1968] that argues individuals revise their beliefs in a manner akin to conservative Bayesians. On the other hand, when subjects are lined up in a circular fashion, the revised prediction the first player does not significantly deviate from the Bayesian benchmark prediction both in the presence of high and low signal.

Comparison of the deviation of the predictions from the Bayesian benchmark prediction shows that, in the linear network setting, the third player makes a less biased prediction compared to the previous players with both high and low signal. Similarly, in the circular network setting, the first player's revised prediction is on average less biased compared to the first and second player's prediction with

[3]There is a significant difference with respect to the second player. The prediction of the second player is only 1.48 higher than the Bayesian benchmark prediction on average. In particular, the second player's prediction significantly deviates from the Bayesian benchmark prediction by 1.72 only under the imperfect information.

[4]The only exception is such that the third player does not significantly underestimate the respective Bayesian benchmark prediction with the high signal.

both high and low signal. Lastly, we compare the deviation from the Bayesian benchmark prediction between the third player's prediction in linear network and first player's revised prediction in circular network setting and find that the difference is not significant with both high and low signal. As the last predictions made in both settings are comparable with each other, we conclude that there is not enough evidence to suggest that the structure of the network plays a role on the predictions.

Next, we ask subjects to solve a series of questions in a Raven Progressive Matrices [Raven et al., 2003] and rank them based on their performance in the test in order to provide them a feedback in the performance context. We ask subjects to predict their own rank and another participant's rank. We use two different setups with respect to the target player's rank: 1^{st} player version and 2^{nd} player version. Accordingly, in the 1^{st} player version, the subjects predict the first player's rank. On the other hand, in the 2^{nd} player's version, the subjects make a prediction about the second player's rank. After eliciting initial beliefs, the first and the second players receive a private signal about the target rank and make a prediction sequentially. The second player will also receive an information about the prediction of the first player. After that, the first player receives an information about the second player's prediction and revises his prediction.

First, we elicit subjects' initial belief about their own rank in the test. We find that subjects significantly overestimate their own rank. This overconfidence is well documented in the literature [Burks et al., 2013; Moore and Healy, 2008; Svenson, 1981]. We also see that the initial belief of the bottom half ranked subjects is more biased compared to the top half ranked subjects as in Gneezy et al. [2017]; Dunning et al. [2003], but we acknowledge that the part of the difference may stem from the ceiling effects for the top half ranked players[5]. Both genders overestimate their rank while the females are significantly less confident with their initial assessment on their own rank than the males controlling for the player's rank in line with Buser

[5]Frequency of the later predictions shows that the top-half ranked subjects overestimate their rank less compared to the bottom-half ranked subjects whereas they underestimate their rank more compared to the bottom-half ranked subjects consistent with the findings by Kruger and Dunning [1999].

et al. [2014]; Barber and Odean [2001].

When we give subjects a private signal and ask them to predict their own rank in the test, the players overestimate their rank. Even though the bias of their predictions about their own rank is relatively less compared to their initial beliefs, the difference between the prediction and the initial belief is not significant. On the other hand, when the subjects receive a private signal and an information about the prediction of the previous player, they again overestimate their own rank in line with the findings of Huffman et al. [2022] that asserts the persistence of the overconfidence with managers in the face of feedback. This time though, the bias of the prediction is significantly less than the bias of the initial belief. Furthermore, when subjects are given an information about the other player's prediction and asked to revise their prediction about their own rank, they still overestimate their rank. Compared to their initial belief and previous prediction, subjects make a less biased choice with their revised prediction. Our results extends the findings of Grossman and Owens [2012] that finds overestimation of one's rank both before and after receiving an information about one's own rank by underlining that the degree of the overestimation decreases with increasing information.

Next, we compare the predictions of the subjects when the object of interest is self relevant to the case of non-relevant. In other words, we compare the scenario where subject predicts his own ranking to the scenario in which subject predicts another participant's rank. Considering only the predictions and excluding the revised predictions, we find that on average the subjects rank themselves higher compared to the other participants. This type of behavior where self-relevance affects subjects to overestimate their performance compared to other parties is well documented in the literature [Ertac, 2011; Grossman and Owens, 2012]. Moreover, we find that the prediction of the self-relevant item is more biased than that of the non-relevant item on average. Lastly, the comparison of the revised predictions reveal that there is no significant difference between the prediction of one's own rank and the other's rank[6]. In fact, subjects do not only overestimate their own rank but

[6]Although we find that the revised prediction of the relevant rank is more biased than the

also other's rank with respect to their revised prediction parallel to the findings of the previous study by Bordalo et al. [2019]. Therefore, we observe that the difference between predicting one's own rank and other player's rank vanishes when the player is asked to revise his prediction. This is in line with the findings of Coutts [2019] that does not find disparities in the later predictions across self-relevant and non-relevant settings.

While the prediction of the subjects do not deviate from the Bayesian benchmark prediction when they are asked to predict a randomly chosen number, their prediction differs significantly from the target rank when subjects predict a player's rank in a Raven Progressive Matrices [Raven et al., 2003] test[7]. Our results extend the findings of Charness et al. [2011]; Eil and Rao [2011]; Ertac [2011]; Möbius et al. [2014] that show differences between the information processing of the ego-relevant and non-relevant object of interest by demonstrating that the predictions becomes biased when the predicted item is self-relevant to at least one party in the network.

The rest of the paper is organized as follows: in section 3.2 we discuss the details of the experiment, in section 3.3 we report our empirical findings, and section 3.4 concludes.

3.2 Experimental design and procedures

In the experiment, a group of individuals sequentially make a prediction about a predetermined number. The goal is to predict a predetermined (target) number that ranges between 0 and 100[8]. First, a random number between 0 and 100 will be drawn by computer, called the target number. Then, the subjects will receive a private signal that will be seen only by the respective subject to estimate a target number.

non-relevant rank, the difference between these two biases are only 0.02.

[7]The only exception is that when the first player predicts the second player's rank in the 2nd player version, there is no significant difference from the second player's actual rank. This is expected to a certain extent since the first player only uses his private signal to make a prediction about the second player's rank. In all other scenarios, either players make a prediction about their own rank or are exposed to the prediction of the target players before making a prediction.

[8]In Section 3.2.4, the objective is to predict the test ranking of one of the subjects in the respective group in which the rank ranges between 1 and 12 with 1 being the best possible rank.

Private signal will truthfully disclose an information about the target number. It will be given as the comparison of the target number to a randomly drawn integer between 0 and 100 (e.g. target number is less than 35, target number is greater than 72). Each subject other than the first subject will also receive an information regarding the prediction of the previous players. This information will be conveyed either as partial (imperfect information) or as full (perfect information). Afterwards, subjects will make a prediction of the target number depending on their signal. The experiment contains five stages and a post-experiment survey.

Table 3.1: Overview of the Treatments in the Experiment

Treatment	Information Type	Player Type	Rounds
1.1: Linear Social Learning	Partial	-	3
1.2: Linear Social Learning	Full	-	3
2.1: Circular Social Learning	Partial	-	2
2.2: Circular Social Learning	Full	-	2
3.1: Circular Social Learning on Test Ranking	Partial	1st Player	2
3.2: Circular Social Learning on Test Ranking	Full	1st Player	2
3.3: Circular Social Learning on Test Ranking	Partial	2nd Player	2
3.4: Circular Social Learning on Test Ranking	Full	2nd Player	2

The overview of the treatments in the experiment can be seen in Table 3.1. Each subject goes through the treatments 1.1, 1.2, 2.1 and 2.2 in Table 3.1 to make a prediction about a target number for a total of 10 rounds. On the other hand, each subject can only take part in one of the four treatments 3.1, 3.2, 3.3 or 3.4 in Table 3.1 to make a prediction about target participant's rank in the test.

3.2.1 Linear Social Learning

In this stage, subjects predict a randomly chosen number by computer called target number for 6 consecutive rounds. Subjects are formed into groups of 3 players indexed with i=1,2,3. In each round, groups will be reformed randomly. Each subject will receive a private signal to predict the target number for the respective round. A

private signal is given as the comparison of the target number to a randomly drawn integer that ranges between 0 and 100. Every player other than the 1st player in each group will also receive an information regarding the decision of the previous players in their group. This information will be partial in the first 3 rounds and full in the last 3 round. If partial information is provided, then the players will be informed that the previous players in their group have made a prediction that is either greater or less than 50 depending on the prediction of the previous players[9]. If full information is provided, then the subsequent players will know the previous players' prediction about the target number.

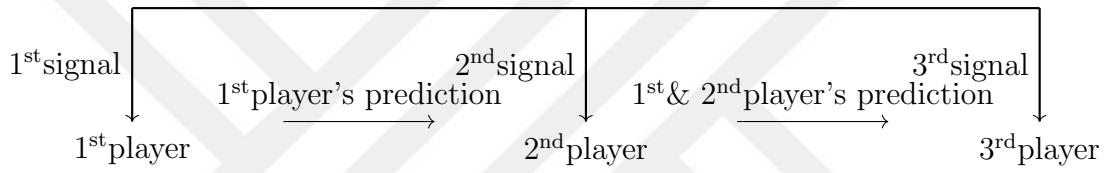


Figure 3.1: Linear Social Learning

The linear social learning stage of the experiment is denoted above in the Figure 3.1. Each subject will assume the role of the 1st, 2nd and 3rd player in both the partial and the full information setting to make a prediction about the target number. For instance, if a subject takes the role of 1st player in the first round and 2nd player in the second round of the partial information setting, then she will be the 3rd player in the last round of the partial information setting.

The payoff from this stage is carried out using a quadratic scoring rule in which the payoff is inversely related to the distance between the subject's prediction and the actual target number. The overall payment from this section is determined by choosing one of the rounds in the partial information and one of the rounds in the full information setting. The formula of the payoff for the payment rounds in tokens from this section is the following:

[9]In partial information round, if a player predicts that the target number is exactly 50, then the partial information conveyed to the subsequent players will be the previous player's prediction is greater than 50 with 0.5 probability and less than 50 with 0.5 probability.

$$100 - (prediction - target\ number)^2/100$$

3.2.2 Circular Social Learning

Different from the first stage, subjects predict a randomly chosen number by computer called target number that ranges between 0 and 100 for only 4 rounds. Subjects are formed into groups of 2 players indexed with i=1,2. In each round, groups will be reformed randomly. Each subject will receive a private signal that will be seen only by the respective subject to predict the target number for the corresponding round. The first 2 round of this stage will be with partial information whereas the last 2 round will be with full information. Similar to Section 3.2.1, the 1st and 2nd players will receive a private signal to make their prediction sequentially. After 2nd player makes a prediction, 1st player will receive an information about the 2nd player's prediction and be asked to revise its prediction.

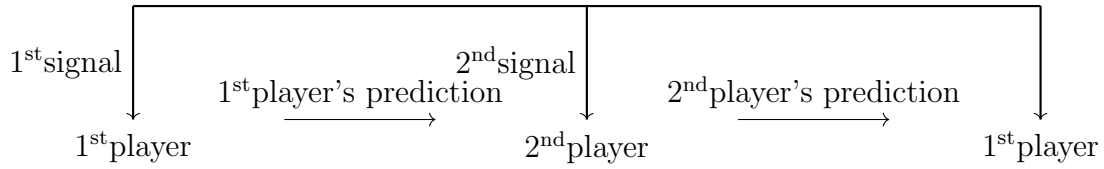


Figure 3.2: Circular Social Learning

The circular social learning stage of the experiment is denoted above in the Figure 3.2. Each subject will assume the role of the 1st and the 2nd player in both the partial and the full information setting to make a prediction about the target number. For example, if a subject takes the role of 1st player in the first round of the full information setting, then she will be the 2nd player in the other round of the full information setting.

The overall payment from this section is determined by choosing one of the rounds in the partial information and one of the rounds in the full information setting. The formula of the payoff for the payment rounds in tokens from this section

is the following:

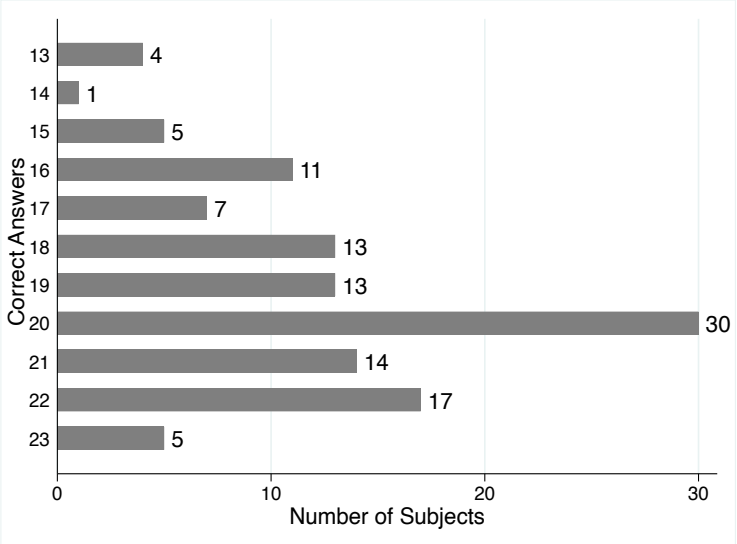
$$100 - (prediction - target\ number)^2/100$$

3.2.3 Raven's Progressive Matrices

We use a version of the Raven's Progressive Matrices test to elicit participant's cognitive ability Raven et al. [2003]. Subjects are asked to solve 23 questions in 10 minutes. While each correct answer is awarded with 10 ECU, wrong answer does not affect the payments from this stage of the experiment. Afterwards, subjects are ranked based on the number of correct answers from the test where ties are broken randomly.

Figure 3.3 demonstrates the distribution of the number of correct answers in Raven's Progressive Matrices. We observe that the number of correct answers range between 13 and 23 (the total number of questions in the test) with 30 subjects having a 20 correct answers out of 23 questions.

Figure 3.3: Distribution of the Number of Correct Answers in RPM



3.2.4 Circular Social Learning on Test Ranking

The prediction of a target number may vary when there is a self relevance of signal. We divide participants into groups of 2 players where players are indexed with i=1,2. The objective is to predict 1^{st} (2^{nd}) player's ranking from Raven's Progressive Matrices test in the previous stage of the experiment.

We use four different versions of this treatment where the differences stem from either the type of information provided or the type of player whom the subjects predict the ranking of. The respective versions are shown in Table 3.2 below:

Table 3.2: Alternative Versions of Rank Prediction

Version	Player Type	Information Type
1^{st} version	1^{st} Player	Partial Information
2^{nd} version	1^{st} Player	Full Information
3^{rd} version	2^{nd} Player	Partial Information
4^{th} version	2^{nd} Player	Full Information

Each subject can only take part in one of the four versions denoted above in Table 3.2. Accordingly, we employ a between-subject design in this stage of the experiment contrary to the within-subject design in the first two stage of the experiment (section 3.2.1 and 3.2.2).

This stage lasts for 2 round where groups are randomly reformed at each round. If a participant takes the role of 1^{st} player in the first round, then the participant becomes the 2^{nd} player in the second round; whereas if a subject is the 2^{nd} player in the first round; then the subject will be the 1^{st} player in the second round. Therefore, each subject will predict her own ranking in one of the rounds and that of an other participant in the other round.

In the 1^{st} Player Type, the target is the rank of the 1^{st} player in the test. The order of steps are as follows:

1. The 1^{st} player states his initial belief about his rank in the test

2. The 1st player receives a private signal in the form of comparison of his ranking to that of a random player in the session (i.e. your rank is better than that of randomly chosen player who has a rank of 8) and makes a prediction

3. The 2nd player receives a private signal about the 1st player's ranking in addition to information about the 1st player's prediction and makes a prediction

4. The 1st player receives an information regarding the 2nd player's prediction and makes a revised prediction

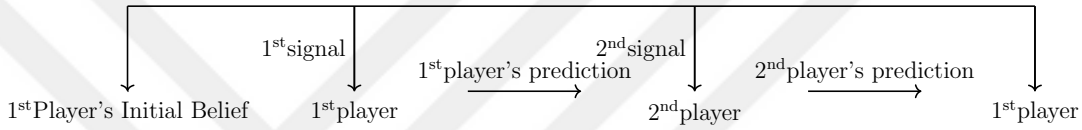


Figure 3.4: 1st Player Type

In the 2nd Player Type, the target is the rank of the 2nd player in the test. The order of steps are as follows:

1. The 2nd player states her initial belief about his rank in the test

2. The 1st player receives a private signal in the form of comparison of 2nd player's ranking to that of a random player in the session and makes a prediction

3. The 2nd player receives a private signal in addition to information about the 1st player's prediction and makes a prediction about her rank in the the test

4. The 1st player receives an information regarding the 2nd player's prediction and makes a revised prediction

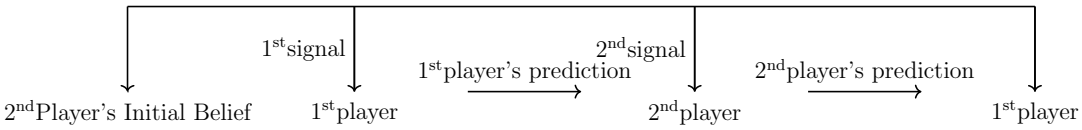


Figure 3.5: 2nd Player Type

The structure of this stage is similar to stage 3.2.2 except with one additional step: we ask the target player to state an initial prediction regarding her own ranking without providing any signal in order to elicit target player's beliefs on her own performance.

Lastly, the overall payment from this section is determined by choosing one of predictions. The formula for the payoff in tokens from this section is the following:

$$150 - (prediction - target\ rank)^2$$

3.2.5 Risk Preferences Elicitation

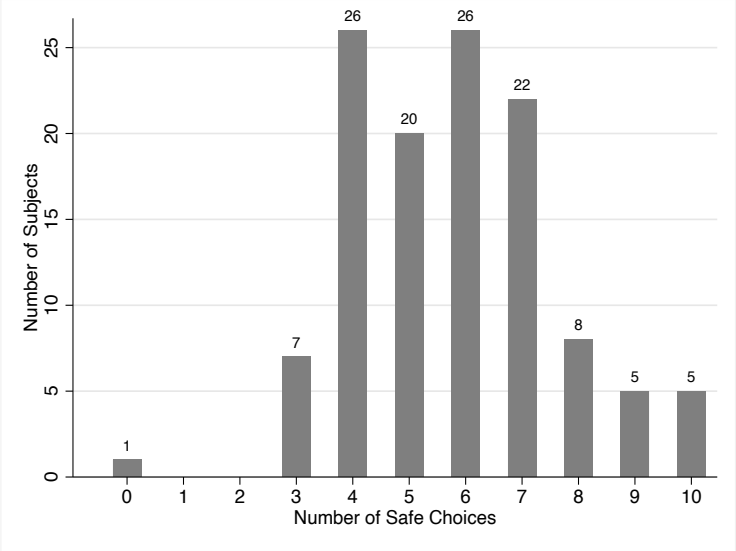
Risk preferences might have an effect on participants' choices in stages 3.2.1, 3.2.2 and 3.2.4 where subjects make a prediction. In order to control for this effect, we elicit risk preferences of each subject at the end of the experiment using a version of the Holt-Laury task [Holt and Laury, 2002]. Subjects are presented with a series of choices between riskier and safer lotteries as shown in Table C.1 in Appendix C.5 and the number of safe choices is taken as a measure of risk aversion.

Figure 3.6 demonstrates the distribution of the number of safe choices in Holt-Laury risk preferences elicitation task.

3.2.6 Experimental procedures

The experiment was conducted with 120 subjects in 2021 from Boğaziçi University with 10 sessions each having 12 subjects. An e-mail was sent to subjects who previously indicated an interest in joining economics experiments, and subjects could register online for a date and time they choose. The experiments were implemented with oTree [Chen et al., 2016] and conducted online. Subjects were informed about the experimental procedure upon their arrival to the Zoom meeting, and the instructions were read aloud. Each subject received a unique URL link which they can use to start to the experiment through a browser on their computer. No subject participated more than once, and the sessions lasted on average 50 minutes. Subjects were paid via electronic fund transfer at the end of the experiment. Subjects received a

Figure 3.6: Distribution of the Number of Safe Choices



participation fee of 10 Turkish Lira in addition to the average payoff of 79.88 Turkish Lira from the experiment, corresponding to a total earnings of 89.88 Turkish Lira on average[10].Random reshuffling of the presentation order on the computer screens ensured that the identity of the players could not be traced over periods.

3.3 Results

We start with the descriptive statistics regarding the behavior of subjects from 3.2.1 and 3.2.2 section of the experiment. Table 3.3 summarizes the mean values and the standard errors of predictions according to the role of the player, separately for the partial and the full information treatments.

[10]At the time of the experiment (08/07/2021- 19/07/2021), 1 Turkish Lira was approximately equal to $0.12 United States dollars.

Table 3.3: Predictions by the role of player

	Imperfect Information			Perfect Information		
	P I	P II	P III	P I	P II	P III
Mean	48.18	48.85	48.38	53.38	55.59	53.45
SE	(22.75)	(18.98)	(20.05)	(23.54)	(22.47)	(23.35)
N	240	240	120	240	240	120

Notes: The table reports the average predictions according to the player role and the type of information in the linear social learning.

3.3.1 1st Player

Let t be the target number and s_1 be the random number to convey the private signal to the first player. If a risk neutral, rational agent receives a private signal, her prediction will be the following:

$$BR_{p_1}(s_1) = s_1/2 \text{ if } s_1 >= t \tag{3.1}$$

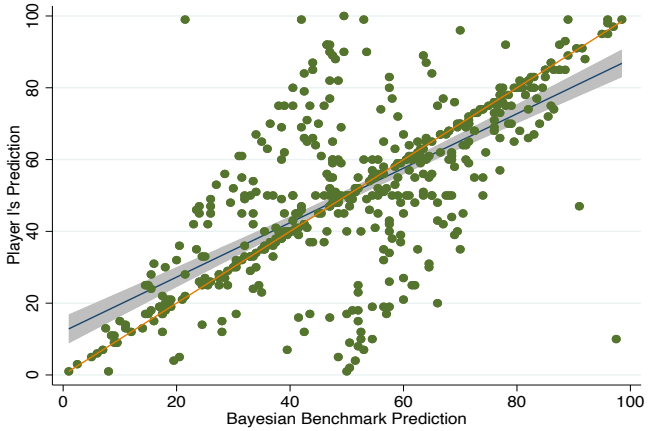
$$BR_{p_1}(s_1) = (s_1 + 100)/2 \text{ if } s_1 < t \tag{3.2}$$

Above in equation 3.1, the target number (t) is less than or equal to the random number (s_1) whereas in equation 3.2 it is greater than the random number. The Bayesian responses to the private signals are given above.

Figure 3.7 demonstrates the predictions of the first player against the Bayesian benchmark predictions. The slope between the first player's prediction and the Bayesian benchmark prediction and is 0.759 (with a standard error of 0.038).

First, we measure the magnitude to which 1st player's prediction deviates from the BB prediction and find that on average 1st player's prediction is 0.23 ($sd = 17.83$) lower than the Bayesian benchmark prediction. Nonetheless, this difference is not

Figure 3.7: Player I's Prediction



Notes: The horizontal axis denotes the prediction made by the 1st player and the vertical axis denotes the Bayesian benchmark prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 1st player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

significant (Wilcoxon matched-pairs signed-rank test, $p = 0.482$)[11]. Overall, this suggests that subjects can incorporate their private signal while making a prediction regarding the target number.

If players are rational, risk-neutral agents, then the Bayesian benchmark prediction would be the Bayesian response to the target number with given information. Accordingly, we show that on average the 1st player underestimates the Bayesian benchmark prediction by 7.84 ($sd = 15.93$) when the target number is greater than or equal to the randomly chosen number (We refer to this type of private signal as *high signal* for the rest of the paper). The underestimation compared to the Bayesian benchmark prediction with high signal is significant (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). On the other hand, the 1st player overestimates

[11]All non-parametric statistical tests reported in this paper are two tailed and take total population, or the respective communities as units of observations. Unless stated otherwise, comparisons within a treatment are tested with the Wilcoxon matched-pairs tests whereas comparisons across treatments are done with the Mann-Whitney tests.

the Bayesian benchmark prediction by 7.91 ($sd = 16.09$) when the signal is such that the target number is less than the random number (We call this type of private signal as *low signal* for the rest of the paper). The overestimation compared to the Bayesian benchmark prediction with low signal is significant (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). There is also a significant difference in the average deviation from the Bayesian benchmark prediction between different signal types (Mann-Whitney test, $p = 0.000$).

Result 1 *The 1st player extracts the information in the private signal to predict the target number. However, she underestimates (overestimates) the Bayesian benchmark prediction when the the target number is larger (smaller) than randomly chosen number in a manner akin to conservative Bayesians.*

3.3.2 2nd Player

Let t be the target number, s_2 be the random number to convey the private signal to the second player and T_{p_1} be the prediction the first player. Assuming that the players are risk-neutral rational agents and they have rational expectations about the other players' decisions, then the second player can decompose the first player's signal via her prediction in the perfect information. Accordingly, the Bayesian responses to the private signals can be differentiated into four different cases depending on the prediction of the first player and the private signal of the second player:

$$BR_{p_2}(s_2) = min(2 * T_{p_1}, s_2)/2 \quad \text{if} \quad T_{p_1} <= 50 \quad \text{and} \quad s_2 >= t \tag{3.3}$$

$$BR_{p_2}(s_2) = (2 * T_{p_1} - 100 + s_2)/2 \quad \text{if} \quad T_{p_1} > 50 \quad \text{and} \quad s_2 >= t \tag{3.4}$$

$$BR_{p_2}(s_2) = (2 * T_{p_1} + s_2)/2$$

$$\text{if } T_{p_1} <= 50 \text{ and } s_2 < t \quad (3.5)$$

$$BR_{p_2}(s_2) = (100 + max(2 * T_{p_1} - 100, s_2))/2$$

$$\text{if } T_{p_1} > 50 \text{ and } s_2 < t \quad (3.6)$$

In the case of imperfect information, if a rational, risk-neutral agent receives a private signal, her prediction will be the following:

$$BR_{p_2}(s_2) = (2 * s_2 * (150 - s_2))/(3 * (200 - s_2))$$

$$\text{if } T_{p_1} <= 50 \text{ and } s_2 >= t \quad (3.7)$$

$$BR_{p_2}(s_2) = 2 * s_2/3$$

$$\text{if } T_{p_1} > 50 \text{ and } s_2 >= t \quad (3.8)$$

$$BR_{p_2}(s_2) = (100 + 2 * s_2)/3$$

$$\text{if } T_{p_1} <= 50 \text{ and } s_2 < t \quad (3.9)$$

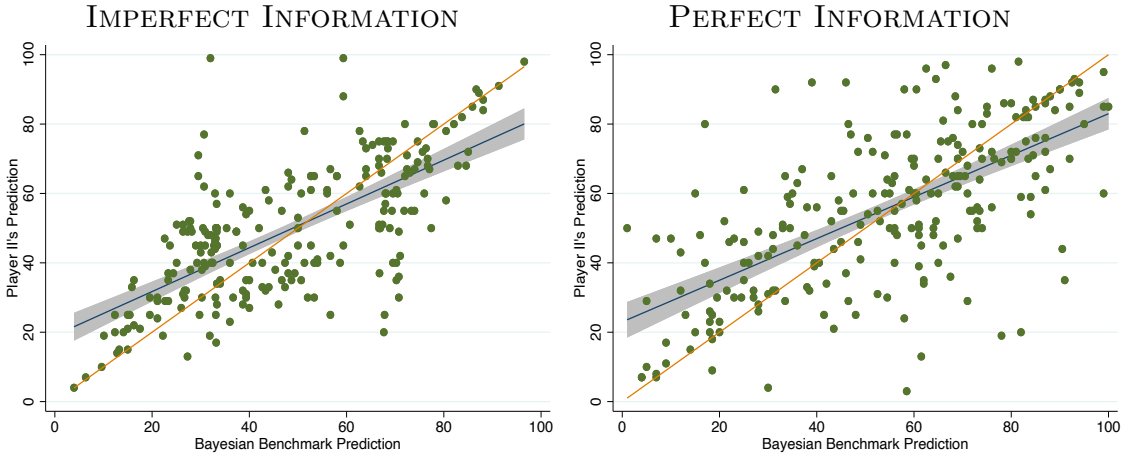
$$BR_{p_2}(s_2) = \frac{(2 * (100 + s_2) * (100 + s_2) - (200 * s_2))}{(3 * (100 + s_2))}$$

$$\text{if } T_{p_1} > 50 \text{ and } s_2 < t \quad (3.10)$$

Figure 3.8 demonstrates the predictions of the second player against the Bayesian benchmark predictions. The slope between the second player's prediction and the

Bayesian benchmark prediction is 0.631 (with a standard error of 0.042) and 0.600 (with a standard error of 0.044) in the case of imperfect information and perfect information, respectively ($p = 0.436$)[12].

Figure 3.8: Player II's Prediction



Notes: The horizontal axis denotes the prediction made by the 2nd player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 2nd player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

The 2nd player's prediction is 1.48 ($sd = 17.72$) higher than the Bayesian benchmark prediction on average (Wilcoxon matched-pairs signed-rank test, $p = 0.030$). While the average deviation of the 2nd player's prediction from the Bayesian benchmark prediction is 1.72 ($sd = 15.68$) under the imperfect information (Wilcoxon matched-pairs signed-rank test, $p = 0.052$), it is on average 1.25 ($sd = 19.58$) under the perfect information (Wilcoxon matched-pairs signed-rank test, $p = 0.253$).

If the 2nd player is rational, risk-neutral agent and have rational expectations about the 1st player, then the Bayesian benchmark prediction would be the Bayesian response to the target number with the provided information. On average, the 2nd player underestimates the Bayesian benchmark prediction by 6.10 ($sd = 16.86$)

[12]When we do not distinguish between information types, the slope between the second player's prediction and the Bayesian benchmark prediction is 0.621 with a standard error of 0.031.

when the target number is greater than or equal to the random number (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). On the contrary, subject overestimates the Bayesian benchmark prediction by 9.01 ($sd = 15.18$) when the signal is such that the target number is less than the random number (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). Moreover, there is a significant difference in the average deviation from the Bayesian benchmark prediction between different signal types (Mann-Whitney test, $p = 0.000$).

The comparison of the deviation of the predictions from the Bayesian benchmark prediction between the 1st and the 2nd players reveals that the deviation is -0.23 ($sd = 17.83$) and 1.48 ($sd = 17.72$) with respect to the first and second player's prediction (Mann-Whitney test, $p = 0.040$).

Lastly, when the players receive a high signal, then both the 1st and the 2nd players underestimate the Bayesian benchmark prediction by 7.84 ($sd = 15.93$) and 6.10 ($sd = 16.86$), respectively (Mann-Whitney test, $p = 0.755$). On the other hand, there is a significant difference in the magnitude of overestimation between the first and second player's prediction with 7.91 ($sd = 16.09$) and 9.01 ($sd = 15.18$) in the presence of low signal (Mann-Whitney test, $p = 0.050$).

Result 2 *The 2nd player underestimates (overestimates) the Bayesian benchmark prediction when the the target number is larger (smaller) than randomly chosen number similar to the 1st player. However, the second player overestimates the Bayesian benchmark prediction more on average compared to the first player with low signal.*

3.3.3 3rd Player

Let t be the target number and s_3 be the random number to convey the private signal to the third player. Assuming that the third player is rational, risk-neutral, her Bayesian response upon receiving a private signal under the perfect information will be the following:

$$BR_{p_3}(s_3) = min(2 * T_{p_1}, 2 * T_{p_2}, s_3)/2$$
$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} <= T_{p_1} \text{ and } s_3 >= t \quad (3.11)$$

$$BR_{p_3}(s_3) = (min(2 * T_{p_1}, 2 * T_{p_1}) + s_3)/2$$
$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} <= T_{p_1} \text{ and } s_3 < t \quad (3.12)$$

$$BR_{p_3}(s_3) = (min(2 * T_{p_1}, s_3) + 2 * (T_{p_2} - T_{p_1}))/2$$
$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > T_{p_1} \text{ and } s_3 >= t \quad (3.13)$$

$$BR_{p_3}(s_3) = (max(2 * (T_{p_2} - T_{p_1}), s_3) + 2 * T_{p_1})/2$$
$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > T_{p_1} \text{ and } s_3 < t \quad (3.14)$$

$$BR_{p_3}(s_3) = (min(2 * (T_{p_2} - T_{p_1}) + 100, s_3) + 2 * T_{p_1} - 100)/2$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} <= T_{p_1} \text{ and } s_3 >= t \quad (3.15)$$

$$BR_{p_3}(s_3) = (max(2 * T_{p_1} - 100, s_3) + 2 * (T_{p_2} - T_{p_1}) + 100)/2$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} <= T_{p_1} \text{ and } s_3 < t \quad (3.16)$$

$$BR_{p_3}(s_3) = (max(2 * T_{p_1} - 100, 2 * T_{p_2} - 100) + s_3)/2$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > T_{p_1} \text{ and } s_3 >= t \quad (3.17)$$

$$BR_{p_3}(s_3) = (100 + max(2 * T_{p_1} - 100, 2 * T_{p_2} - 100, s_3))/2$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > T_{p_1} \text{ and } s_3 < t \quad (3.18)$$

In the imperfect information case, the number of scenarios are multi-fold with respect to the private signal and the partial information regarding the prediction of the first and second player. The Bayesian response of the third player under imperfect information assuming that the player is a risk-neutral, rational agent can be found in the C.2 section of the appendix.

Figure 3.9 demonstrates the actual predictions of the third player against the Bayesian benchmark predictions. The slope between the actual prediction and the Bayesian benchmark prediction is 0.683 (with a standard error of 0.071) and 0.855 (with a standard error of 0.052) in imperfect information and perfect information, respectively ($p = 1.000$)[13].

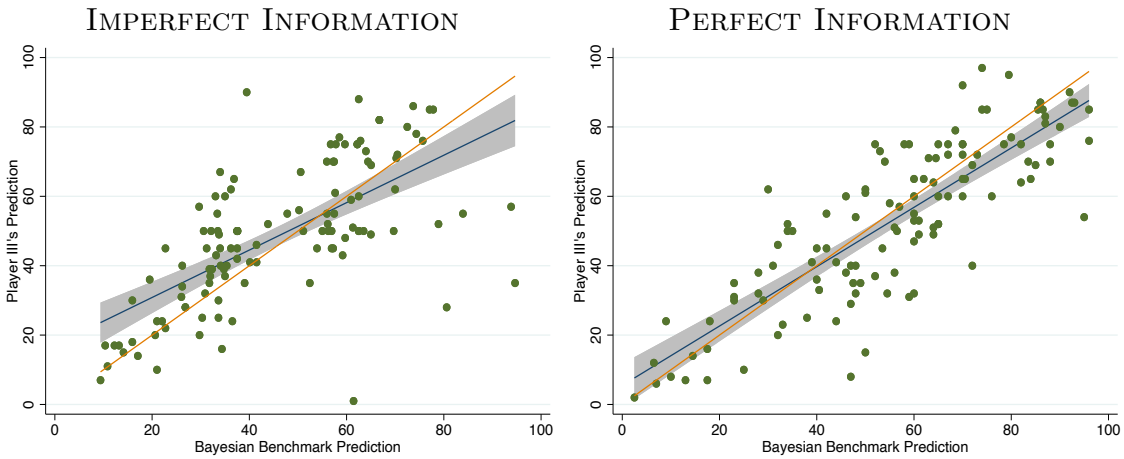
We start with calculating the degree to which 3rd player's prediction deviates from the Bayesian benchmark prediction and find that on average 3rd player's prediction is 0.06 ($sd = 15.04$) higher than the Bayesian benchmark prediction but not statistically significant (Wilcoxon matched-pairs signed-rank test, $p = 0.434$).

Next, we find that the prediction under the imperfect information is on average 2.71 ($sd = 16.22$) higher than the Bayesian benchmark prediction (Wilcoxon matched-pairs signed-rank test, $p = 0.002$). On the other hand, the 3rd player's prediction under the perfect information is on average 2.60 ($sd = 13.31$) lower than the Bayesian benchmark (Wilcoxon matched-pairs signed-rank test, $p = 0.047$).

Assuming that the 3rd player is rational, risk-neutral agent and have rational expectations about the 1st and 2nd player, the Bayesian benchmark prediction would be the Bayesian response to the target number. By this logic, we show that on average the 3rd player underestimates the Bayesian benchmark prediction by 0.88 ($sd = 14.51$) when the target number is greater than or equal to the randomly

[13]When we do not distinguish between information types, the slope between the third player's prediction and the Bayesian benchmark prediction is 0.766 with a standard error of 0.042.

Figure 3.9: Player III's Prediction



Notes: The horizontal axis denotes the actual prediction made by the 3rd player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the third player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

chosen number (Wilcoxon matched-pairs signed-rank test, $p = 0.510$). On the other hand, subject overestimates the Bayesian benchmark prediction by 1.02 (Wilcoxon matched-pairs signed-rank test, $sd = 15.58$) when the signal is such that the target number is less than the random number ($p = 0.058$). There is a marginal significant difference in the average deviation from the Bayesian benchmark prediction between different signal types (Mann-Whitney test, $p = 0.080$).

When it comes to the comparison of the deviation of the predictions from the Bayesian benchmark prediction between the 1st and the 3rd players, the deviation is -0.23 ($sd = 17.83$) and 0.06 ($sd = 15.04$) with respect to the first and third player's prediction (Mann-Whitney test, $p = 0.326$). On the other hand, the 1st and the 3rd players underestimate the Bayesian benchmark prediction by 7.84 ($sd = 15.93$) and 0.88 ($sd = 14.51$) with high signal (Mann-Whitney test, $p = 0.001$), and overestimate by 7.91 ($sd = 16.09$) and 1.02 ($sd = 15.58$) with low signal (Mann-Whitney test, $p = 0.027$).

We also compare the deviation of the predictions from the Bayesian benchmark

prediction between the 2nd and the 3rd players, and find that the deviation is 1.48 ($sd = 17.72$) and 0.06 ($sd = 15.04$) with respect to the second and third player's prediction (Mann-Whitney test, $p = 0.406$). While the difference between these two players is insignificant with the average deviation of 0.77 ($sd = 19.15$) and 1.72 ($sd = 15.68$) under the imperfect information (Mann-Whitney test, $p = 0.285$), it is significant under the perfect information with the average deviation of 1.25 ($sd = 19.58$) and -2.60 ($sd = 13.31$) for the second and third players (Mann-Whitney test, $p = 0.031$). Finally, the 2nd and the 3rd players underestimate the Bayesian benchmark prediction by 6.10 ($sd = 16.86$) and 0.88 ($sd = 14.51$) with high signal (Mann-Whitney test, $p = 0.001$), and overestimate by 9.01 ($sd = 15.18$) and 1.02 ($sd = 15.58$) with low signal (Mann-Whitney test, $p = 0.000$).

Lastly, we run a series of OLS regressions where the dependent variable is the deviation of the player's prediction from the corresponding Bayesian benchmark prediction in Table 3.4[14]. The independent variable, high signal is a binary variable that takes the value of 1 with high signal and 0 with low signal. The information level is equal to 1 with perfect information scheme and 0 otherwise. The independent variables, 1st and 2nd player's prediction takes the value of 1 if the respective prediction is greater than or equal to 50 and 0 otherwise. Lastly, female is binary variable that is equivalent to 1 if the player is female and 0 if male[15]. We also control for certain attitudinal and demographic information about subjects[16].

While the signal type has a significant effect on the deviation of the prediction from the Bayesian benchmark prediction for the 1st and 2nd players, it does not have a significant effect for the 3rd player. The significance of this variable is robust to controlling factors for the first two players. The first player's prediction has a significant effect on the deviation of the subsequent players prediction. This effect

[14]The deviation is calculated as the player's prediction minus the corresponding Bayesian benchmark prediction

[15]1 out of 120 subject did not fill out the gender part of the survey

[16]The set of all control variables used in regressions are as follows: age in years, a gender dummy, year of study in the degree, major (1 if quantitative, 0 otherwise), number of correct answers to 5 math problems in the survey (Section C.5.6), and the number of safe choices in Holt-Laury task [Holt and Laury, 2002].

is significant controlling for other factors (Model 4 and 6). On the other hand, the second player's prediction does not have a significant effect on the deviation of the third player's prediction. The information level has a significant impact only on the deviation of the third player's prediction and this effect is robust to controlling factors (Model 6). Finally, females overestimate the Bayesian benchmark prediction compared to males only when they are assigned as the third player (Model 6).

Table 3.4: Determinants of the Deviation from Bayesian benchmark

	(1)	(2)	(3)	(4)	(5)	(6)
Deviation	1st Player	1st Player	2nd Player	2nd Player	3rd Player	3rd Player
High Signal	-15.75***	-15.88***	-12.91***	-12.78***	1.580	1.645
	(2.159)	(2.177)	(1.406)	(1.312)	(2.869)	(2.597)
Information Level			1.140	0.947	-4.168**	-4.424**
			(2.186)	(2.057)	(1.556)	(1.508)
1st Player's Prediction			-10.45***	-10.62***	-3.286*	-3.797**
			(1.199)	(1.176)	(1.697)	(1.547)
2nd Player's Prediction					-5.262	-4.406
					(3.366)	(3.308)
Female		-0.178		-0.748		5.594**
		(1.391)		(1.628)		(2.314)
Constant	7.907***	14.81**	13.07***	4.742	6.015***	-4.011
	(1.563)	(5.901)	(0.939)	(7.672)	(1.681)	(17.70)
Controls	No	Yes	No	Yes	No	Yes
N	480	476	480	476	240	238
R^2	0.194	0.204	0.260	0.257	0.063	0.077

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Result 3 *While the 3rd player overestimates the Bayesian benchmark prediction when the the target number is smaller than randomly chosen number similar to the 1st and 2nd player, she does not make a prediction significantly different from the Bayesian benchmark prediction when the target number is greater than randomly*

chosen number. Furthermore, the 3rd player's prediction is on average less biased compared to the 1st and the 2nd player with both high and low signal.

3.3.4 1st Player's Revised Prediction

Let t be the target number and s_1 be the random number to convey the private signal to the first player. Assuming that the first player is rational, risk-neutral, her Bayesian response upon receiving a full information regarding the second player's prediction will be the following:

$$BR_{p_1}^{revised}(s_1) = T_{p_2} \tag{3.19}$$

Under the imperfect information scheme, the first player's revised prediction will coincide with the Bayesian benchmark prediction assuming that the first player is rational, risk-neutral agent. The Bayesian response of the first player upon receiving a partial information regarding the second player's prediction will be the following:

$$BR_{p_1}^{revised}(s_1) = (2 * s_1 + 75)/3$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} <= 50 \text{ and } s_1 < t \tag{3.20}$$

$$BR_{p_1}^{revised}(s_1) = (2 * s_1 + 25)/3$$
$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > 50 \text{ and } s_1 >= t \tag{3.21}$$

$$BR_{p_1}^{revised}(s_1) = (s_1 + 100)/3$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > 50 \text{ and } 75 < s_1 < t \tag{3.22}$$

$$BR_{p_1}^{revised}(s_1) = \frac{((100 * (100 - s_1) * (100 + s_1)/3) - (75 - s_1)^2 * (75 + 2 * s_1)/6)}{((100 - s_1) * 100 - (75 - s_1)^2/2)}$$
$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > 50 \text{ and } s_1 < t \text{ and } s_1 <= 75 \tag{3.23}$$

$$BR_{p_1}^{revised}(s_1) = (s_1)/2$$
$$\text{if } T_{p_1} < 50 \text{ and } T_{p_2} < 50 \text{ and } 25 >= s_1 >= t \quad (3.24)$$

$$BR_{p_1}^{revised}(s_1) = \frac{((100 * s_1 * s_1/2) - ((s_1 - 25) * (s_1 - 25) * (2 * s_1 + 25)/6))}{(100 * s_1 - (s_1 - 25) * (s_1 - 25)/2)}$$
$$\text{if } T_{p_1} < 50 \text{ and } T_{p_2} < 50 \text{ and } s_1 >= t \text{ and } s_1 > 25 \quad (3.25)$$

Figure 3.10 demonstrates the revised predictions of the first player against the Bayesian benchmark predictions. The slope between the Nash prediction and the actual prediction is 0.864 (with a standard error of 0.050) and 0.600 (with a standard error of 0.058) in imperfect information and perfect information, respectively ($p = 0.036$)[17].

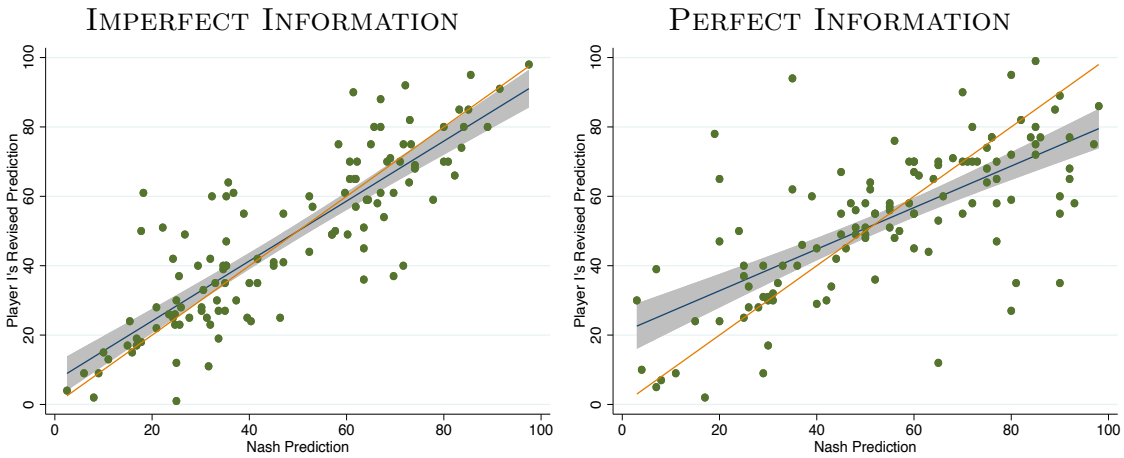
First, we compute the magnitude to which 1st player's revised prediction deviates from the Bayesian benchmark prediction and find that on average it is 0.84 ($sd = 21.69$) lower than the Bayesian benchmark prediction, but not statistically significant (Wilcoxon matched-pairs signed-rank test, $p = 0.947$).

Second, we find that the deviation is on average 0.20 ($sd = 17.51$) under the imperfect information (Wilcoxon matched-pairs signed-rank test, $p = 0.583$). There is also no significant deviation under the perfect information with 2nd player's prediction is on average 1.87 ($sd = 25.23$) lower than the Bayesian benchmark (Wilcoxon matched-pairs signed-rank test, $p = 0.689$).

Third, the Bayesian benchmark prediction would be the Bayesian response to the target number given that the first player is rational, risk neutral agent and have rational expectations about the second player. Accordingly, the 1st player underestimates the Bayesian benchmark prediction by 0.46 ($sd = 23.27$) when the target number is greater than or equal to the randomly chosen number (Wilcoxon matched-pairs signed-rank test, $p = 0.818$). Subject also underestimates the Bayesian benchmark

[17]When we do not distinguish between information types, the slope between the first player's revised prediction and the Bayesian benchmark prediction is 0.727 with a standard error of 0.039.

Figure 3.10: Player I's Revised Prediction



Notes: The horizontal axis denotes the revised prediction made by the 1st player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the first player's revised prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

prediction by 1.26 ($sd = 19.84$) when the signal is such that the target number is less than the random number (Wilcoxon matched-pairs signed-rank test, $p = 0.488$). There is not a significant difference in the average deviation from the Bayesian benchmark prediction between different signal types (Mann-Whitney test, $p = 0.562$).

Table 3.5 shows the mean deviation of the player's prediction from the corresponding Bayesian benchmark prediction. When it comes to the comparison of the deviation of the predictions from the Bayesian benchmark prediction, there is not a significant difference between the first player's prediction and his revised prediction (Wilcoxon matched-pairs signed-rank test, $p = 0.738$). However, there is a significant difference in the deviation with high signal (Wilcoxon matched-pairs signed-rank test, $p = 0.000$) and with low signal (Wilcoxon matched-pairs signed-rank test, $p = 0.007$).

Next, we compare the deviation from the Bayesian benchmark prediction between the 1st player's revised prediction and the 2nd players prediction and find that the difference is not significant (Mann-Whitney test, $p = 0.268$). Neither the dif-

Table 3.5: Mean Deviation from the Bayesian benchmark Prediction

	All		Imperfect Information		Perfect Information		Low Signal		High Signal	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1st Player's Prediction	-0.23	17.83	-	-	-	-	7.91	16.09	-7.84	15.93
2nd Player's Prediction	1.48	17.72	1.72	15.68	1.25	19.58	9.01	15.18	-6.10	16.86
3rd Player's Prediction	0.06	15.04	2.71	16.22	-2.60	13.31	1.02	15.58	-0.88	14.51
1st Player's Revised Prediction	-0.84	21.69	0.20	17.51	-1.87	25.23	-1.26	19.84	-0.46	23.27

Notes: The table reports the average deviation of the predictions from the Bayesian benchmark prediction for the 1st,2nd and 3rd players in the linear social learning in section 3.2.1 and circular social learning in section 3.2.2.

ference in the average deviation between these two predictions under the imperfect information (Mann-Whitney test, $p = 0.417$) nor the difference under the perfect information (Mann-Whitney test, $p = 0.364$) is significant. On the other hand, there is a significant difference both in the magnitude of the deviation with the high signal (Mann-Whitney test, $p = 0.002$), and in the magnitude of the deviation with the low signal (Mann-Whitney test, $p = 0.000$).

We also compare the deviation of the predictions from the Bayesian benchmark prediction between the 1st player's revised prediction and the 3rd player's prediction, and find that the deviation is not significant (Mann-Whitney test, $p = 0.737$). There is a marginal significant difference in the deviation under the imperfect information (Mann-Whitney test, $p = 0.075$); however, there is not a significant difference under the perfect information (Mann-Whitney test, $p = 0.381$). The difference between the deviations is not significant both with high signal (Mann-Whitney test, $p = 0.747$), and low signal (Mann-Whitney test, $p = 0.399$).

Lastly, we run a series of OLS regressions in Table 3.4 where the dependent variable is the deviation of the player's prediction from the Bayesian benchmark prediction for Model 1 and 2, and the subtraction of the 1st player's prediction from his revised prediction for Model 3 and 4[18].

[18]The deviation is calculated as the player's prediction minus the corresponding Bayesian benchmark prediction whereas the update is computed as the player's revised prediction minus the

The signal type, the previous prediction of the first player and the second player's prediction have a significant effect on the deviation of the revised prediction from the Bayesian benchmark prediction (Model 1). The significance of these variables are robust to controlling factors (Model 2). The information level or gender does not have any significant effect on the deviation.

When it comes to the degree of revision, the previous prediction of the first player and the prediction of the second player have a significant impact on the extent of the update (Model 3). The significance of these variables is robust to additional controls (Model 4). Neither the information level nor the gender have any significance on the extent of the update.

Result 4 *The 1st player's revised prediction does not significantly deviate from the Bayesian benchmark prediction in the presence of high (low) signal contrary to the 1st, 2nd and 3rd players regarding their prediction. Moreover, the 1st player's revised prediction is on average less biased compared to the 1st and 2nd player's prediction with both high and low signal.*

3.3.5 Initial Belief and Prediction on Rank

In this section, we analyze the subjects' predictions with respect to ranking of their performance in the Raven's Progressive Matrices (RPM). Each session constitutes 12 players, accordingly players' rank in RPM range between 1 and 12 with 1 being the best possible rank. Below, we report some descriptive statistics with respect to the type of information and the relevancy of the target rank in Table 3.7.

Initial Belief

Figure 3.11 shows the distribution of choices with respect to the beliefs of the players regarding their own performance in Raven Progressive Matrices. Players have an initial belief on their rank with an average value of 4.60 ($sd = 2.26$).

player's previous prediction

Table 3.6: 1st Player's Revised Prediction

	(1) Deviation	(2) Deviation	(3) Update	(4) Update
High Signal	-5.264***	-4.665**		
	(1.595)	(1.807)		
Previous Prediction	0.719***	0.704***	-0.357***	-0.355***
	(0.069)	(0.067)	(0.036)	(0.036)
Information Level	-2.206	-2.195	0.733	0.826
	(2.221)	(2.339)	(1.209)	(1.289)
2nd Player's Prediction	-28.50***	-27.90***	15.15***	15.10***
	(3.008)	(2.721)	(1.665)	(1.634)
Female		1.221		-0.731
		(2.419)		(0.972)
Constant	-17.29***	-39.26***	9.668**	23.82**
	(3.229)	(11.70)	(2.992)	(8.444)
Controls	No	Yes	No	Yes
N	240	238	240	238
R^2	0.565	0.561	0.356	0.355

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

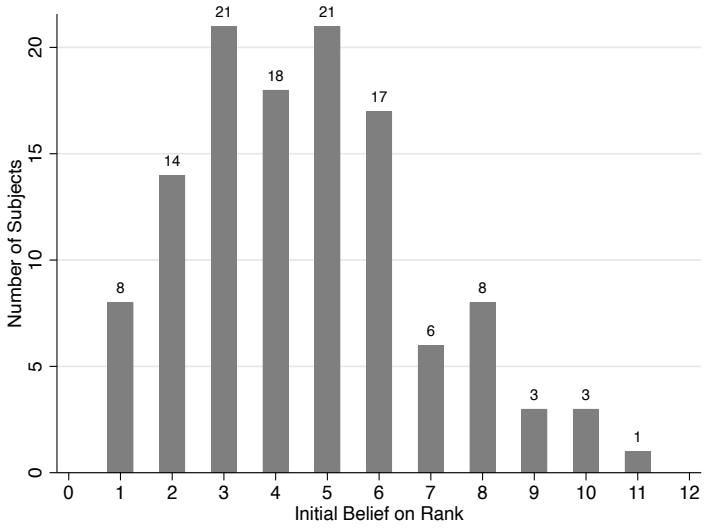
Table 3.7: Predictions by the role of player

	Imperfect Information						Perfect Information					
	Own Rank			Others' Rank			Own Rank			Others' Rank		
	Initial Belief	Prediction	Revised Prediction	Initial Belief	Prediction	Revised Prediction	Initial Belief	Prediction	Revised Prediction	Initial Belief	Prediction	Revised Prediction
Mean	4.90	5.37	5.90	-	6.43	5.10	4.30	4.87	5.17	-	6.52	6.00
SE	(2.36)	(2.26)	(2.52)	(-)	(2.59)	(2.52)	(2.13)	(2.78)	(2.63)	(-)	(2.59)	(2.91)
N	60	60	30	-	60	30	60	60	30	-	60	30

Notes: The table reports the average predictions according to the player role and the type of information in the circular social learning on test ranking.

While 83% (99) of all subjects rank themselves between 1st and 6th places, only 17% (21) of all subjects have an initial belief between 7th and 12th places regarding their own rank as opposed to the expected 50% (test for proportions, $p = 0.000$) pointing towards high prevalence of initial optimism[19].

Figure 3.11: Initial Belief on Rank



To be able to assess whether there is a bias in the initial beliefs, we need to compare the distribution of the initial beliefs to that of the expected distribution with full information. The distribution of the initial beliefs would be a discrete uniform distribution if subjects had full information with respect to their ranking[20]. On the other hand, as it can be seen in Figure 3.11, the distribution of the initial beliefs is skewed to the left and significantly different from a uniform distribution (chi-square test, $p = 0.000$). Initial beliefs have a mean rank of 4.60, instead of the expected 6.50 under the discrete uniform distribution.

Next, we assess the degree to which subjects' initial belief deviates from their actual ranking in the test and find that on average subjects rank themselves as 1.90 ($sd = 3.37$) better than their actual rank. This difference is statistically significant

[19]No player ranked itself on the 12th place.

[20]The respective discrete uniform distribution would range between 1 and 12.

(Wilcoxon matched-pairs signed-rank test, $p = 0.000$). Overall, 64% of the participants (77 subjects) have an initial belief that they are ranked higher than their actual rank. Those individuals' initial belief on average deviate by 3.88 ($sd = 2.29$) ranks from their actual rank.

While only 27% (16) of bottom-half ranked subjects initially predict to be in the bottom-half rank, 92% (55) of top-half ranked subjects place themselves into the top-half rank with respect to their initial beliefs. When we evaluate the extent to which subjects' initial belief deviates from their actual ranking, we find that on average the initial belief of the bottom half ranked subjects is more biased with a mean value of -4.25 ($sd = 2.62$) than that of the top half ranked subjects with a mean value of 0.45 ($sd = 2.19$) and the difference is significant (Mann-Whitney test, $p = 0.000$)[21].

We notice that both genders overestimate their rank. On average, male subjects rank themselves as 2.21 ($sd = 3.25$) better than their actual rank (Wilcoxon matched-pairs signed-rank test, $p = 0.000$), and female subjects rank themselves as 1.62 ($sd = 3.50$) better than their actual rank (Wilcoxon matched-pairs signed-rank test, $p = 0.001$). We find no significant difference in the average deviation between sexes (Mann-Whitney test, $p = 0.361$). Overall, 70% (39) of males rank themselves better than their actual rank with average deviation of 3.77 ($sd = 2.41$) while 59% (37) of females rank themselves better than their actual rank with average deviation of 4.05 ($sd = 2.20$).

Result 5 *The initial beliefs of subjects reveal that there is a bias compared to their actual rank. Also, the bottom half ranked subjects are more biased compared to the top half ranked participants.*

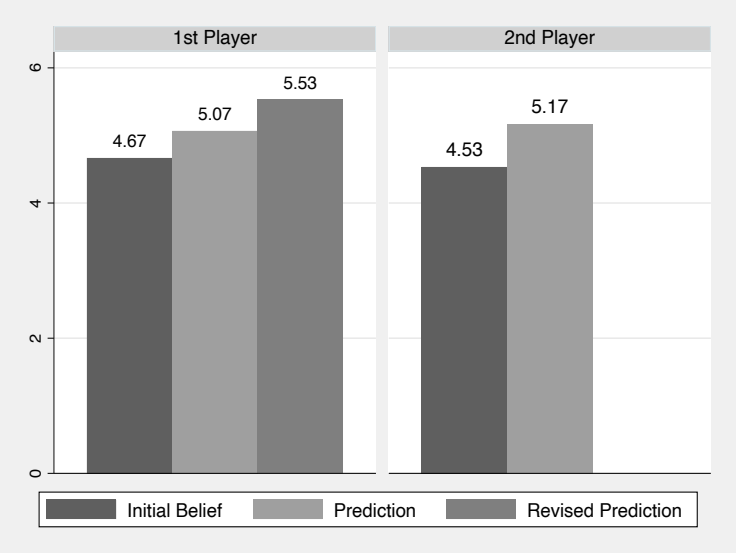
Initial Belief versus Prediction

In this section, we compare the initial belief of the players with their prediction and revised prediction regarding their own rank. Figure 3.12 shows the mean prediction

[21]Even though the difference in deviations is substantial, part of this difference stems from the ceiling effects for the top half ranked players.

of the players on their own rank for the 1st and 2nd player version, respectively.

Figure 3.12: Prediction on Own Rank



For the 1st player version, the players' prediction about their own rank decreases upon receiving private signal with respect to their ranking compared to their initial belief: players have an initial belief on their rank with a value of 4.67 ($sd = 2.11$) whereas their prediction after receiving a private signal worsens to 5.07 ($sd = 2.55$) on average (Wilcoxon matched-pairs signed-rank test, $p = 0.337$).

When it comes to the comparison of the players' initial belief to their revised prediction for the 1st player version, players decrease their assessment by 0.86 to a mean value of 5.53 ($sd = 2.58$) with respect to their own rank (Wilcoxon matched-pairs signed-rank test, $p = 0.008$). There is a marginal significant difference in imperfect information case: players on average have an initial belief on their rank with a value of 5.13 ($sd = 2.22$) whereas their revised prediction decreases to 5.90 ($sd = 2.52$) upon receiving information on the prediction of the second player (Wilcoxon matched-pairs signed-rank test, $p = 0.053$). There is also a marginal significant difference in perfect information case: players have an initial belief on their rank with 4.20 ($sd = 1.92$) whereas their revised prediction worsens to 5.17 ($sd = 2.63$) on average (Wilcoxon matched-pairs signed-rank test, $p = 0.065$).

The comparison of the players' prediction to their revised prediction on their own rank reveals that players revised prediction worsens by 0.46 on average (Wilcoxon matched-pairs signed-rank test, $p = 0.001$). There is a marginal significant difference between these two prediction in the imperfect information case: players on average have a prediction on their rank with 5.50 ($sd = 2.29$) whereas their revised prediction slightly worsens to 5.90 ($sd = 2.52$) upon receiving information about the prediction of the second player (Wilcoxon matched-pairs signed-rank test, $p = 0.058$). In the perfect information case, there is a significant difference as well: players have a prediction on their rank with 4.63 ($sd = 2.76$) whereas their revised prediction worsens to 5.17 ($sd = 2.63$) on average upon receiving information about the prediction of the second player (Wilcoxon matched-pairs signed-rank test, $p = 0.009$).

Next, we compare the initial belief of the players with their prediction regarding their own rank for the 2$^{\text{nd}}$ player version. While players on average have an initial belief of 4.53 ($sd = 2.42$), their prediction decreases to 5.17 ($sd = 2.55$) with respect to their own rank (Wilcoxon matched-pairs signed-rank test, $p = 0.011$). In imperfect information case, there is a marginal significant difference: players on average have an initial belief with 4.67 ($sd = 2.51$) whereas their prediction after receiving a private signal worsens to 5.23 ($sd = 2.27$) concerning their rank (Wilcoxon matched-pairs signed-rank test, $p = 0.061$). In perfect information case, there is a marginal significant difference as well: players on average have an initial belief on their rank with 4.40 ($sd = 2.36$) whereas their prediction decreases to 5.10 ($sd = 2.83$) after receiving a private signal (Wilcoxon matched-pairs signed-rank test, $p = 0.085$).

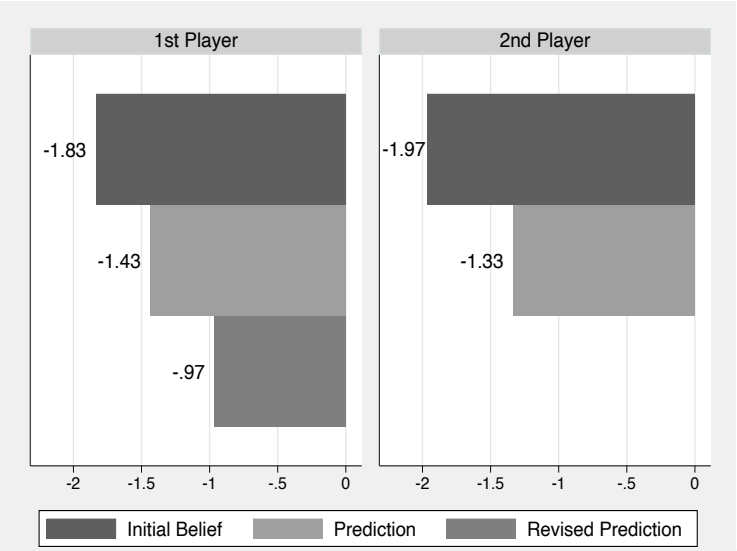
Result 6 *The 1st player's revised prediction on his own rank worsens significantly compared to his previous prediction and his initial belief. Furthermore, the 2nd player's prediction on her own rank worsens compared to her initial belief.*

Prediction versus Actual Rank

In this section, we compare the prediction and revised prediction of the player's regarding their own rank to their actual rank. Figure 3.13 shows the mean deviation

of the players prediction about their rank from their actual rank for the 1st and 2nd player version, respectively.

Figure 3.13: Mean Deviation from Player's Own Rank



We start with the comparison of the 1st player's prediction about his own rank to the respective player's actual rank, and discover that on average the player overestimates his own rank by 1.43 ($sd = 2.19$) and the difference in estimation is significant (Wilcoxon matched-pairs signed-rank test, $p = 0.000$).

When it comes to the deviation of the 1st player's revised prediction from the actual rank of the subject, we see that the player overestimates his own rank by 0.97 ($sd = 1.99$) on average (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). Particularly, the subjects significantly overestimate their rank by 1.33 ($sd = 2.14$) in the imperfect information setting (Wilcoxon matched-pairs signed-rank test, $p = 0.003$), and by 0.60 ($sd = 1.79$) in the perfect information setting on average (Wilcoxon matched-pairs signed-rank test, $p = 0.057$).

Next, we calculate the deviation of 2nd player's prediction from the actual rank of the subject, and find that the player overestimates his own rank by 1.33 ($sd = 2.35$) on average (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). There is a significant difference in both the imperfect and perfect information case: players on

average overestimate their rank by 1.13 ($sd = 2.27$) in the former setting (Wilcoxon matched-pairs signed-rank test, $p = 0.025$), and by 1.53 ($sd = 2.45$) in the latter scenario (Wilcoxon matched-pairs signed-rank test, $p = 0.003$).

Table 3.8 shows the frequency of the cases where the prediction of the subjects is higher, same or lower than their actual rank.

Table 3.8: Comparison of the Prediction to the Actual Rank

Breakdown	Group	PI's Prediction			PI's Revised Prediction			PII's Prediction		
		Higher	Same	Lower	Higher	Same	Lower	Higher	Same	Lower
Actual Rank	Top Half	14	6	10	11	6	13	11	9	10
	Bottom Half	29	1	0	26	3	1	28	0	2
Gender	Female	16	3	5	13	5	6	24	5	10
	Male	27	4	5	24	4	8	14	4	2
	Total	43	7	10	37	9	14	39	9	12

Notes: The table reports the frequency of the prediction by subjects against the actual rank of the subject.

While the prediction of subjects are better than their actual rank 66% (119 out of 180) of all cases, their prediction are either equal or worse than their own rank for the remaining 34% (61 out of 180) cases.

The top half ranked subjects overestimate their rank in only 40% (36 out of 90) of all predictions while the bottom half ranked subjects overestimate it in 92% (83 out of 90) of all available cases (test of proportions, $p = 0.000$). On the other hand, the top half ranked subjects underestimate their own rank 36% (32 out of 90) of all cases whereas the bottom half ranked subjects do so only in 3% (3 out of 90) of all possible cases (test of proportions, $p = 0.000$).

When it comes to the frequency of the genders, female subjects overestimate their rank in 61% (53 out of 87) of all predictions while males overestimate it in 71% (65 out of 92) of all available cases (test of proportions, $p = 0.170$). On the other hand, females underestimate their own rank 23% (21 out of 87) of all cases whereas males subjects do so only in 16% (15 out of 92) of all possible cases (test of proportions, $p = 0.191$).

Lastly, we run a series of OLS regressions in Table 3.9 where the dependent

variable is the deviation of the players' prediction from their own actual rank. The positive values of the dependent variable indicate that subjects underestimate their own rank whereas negative values show that subjects overestimate their own rank.

First, players with lower actual rank overestimates their rank in comparison with the players with higher actual rank in the test. This effect is robust to controlling factors. Second, the signal type has a significant effect on the deviation of the player's prediction from her own rank. In particular, we see that player's that receive high signal overestimates their rank compared to a player that receive low signal. However, this effect is not robust to additional controls (except in Model 8). The initial belief of the players only have a significant effect on the deviation of the 2nd player's later prediction from her own rank[22]. The level of information and the other player's prediction do not have any significant effect on the deviation. While females are more conservative with their initial assessment compared to males (Model 2), later predictions reveal no difference between genders.

Result 7 *Even though the players decrease their predictions about their own rank with new piece of information, the players still overestimate their rank in both the imperfect and perfect information setting. Furthermore, top-half ranked subjects overestimate (underestimate) their rank less (more) compared to the bottom-half ranked subjects.*

[22]The second player with lower initial belief underestimates her own rank when she receives a private signal compared to a second player with relatively more optimistic initial belief.

Table 3.9: Deviation from Own Rank

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deviation	All Players	All Players	1st Player	1st Player	2nd Player	2nd Player	1st Player	1st Player
High Signal			-1.190*	-1.226	-1.427**	-1.332	-0.889**	-0.914*
			(0.526)	(0.618)	(0.427)	(0.731)	(0.310)	(0.383)
Player's Rank	-0.759***	-0.788***	-0.591***	-0.624***	-0.788***	-0.812***	-0.602***	-0.585**
	(0.063)	(0.069)	(0.035)	(0.057)	(0.064)	(0.076)	(0.123)	(0.166)
Initial Belief			0.270	0.241	0.442**	0.395**	0.161	0.134
			(0.133)	(0.130)	(0.104)	(0.112)	(0.187)	(0.193)
1st Player's Prediction					-0.768	-0.960*		
					(0.510)	(0.416)		
2nd Player's Prediction							-1.042	-0.977
							(0.862)	(1.068)
Information Level					0.111	-0.0912	0.283	0.315
					(0.242)	(0.266)	(0.214)	(0.274)
Female		1.068*		0.204		0.670		0.362
		(0.479)		(0.295)		(0.511)		(0.368)
Constant	3.036***	3.223	1.742	1.287	2.857***	1.134	2.967	2.845
	(0.330)	(2.422)	(0.904)	(2.499)	(0.510)	(1.194)	(1.795)	(3.438)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	120	119	60	60	60	59	60	60
R^2	0.607	0.620	0.579	0.558	0.683	0.683	0.549	0.521

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

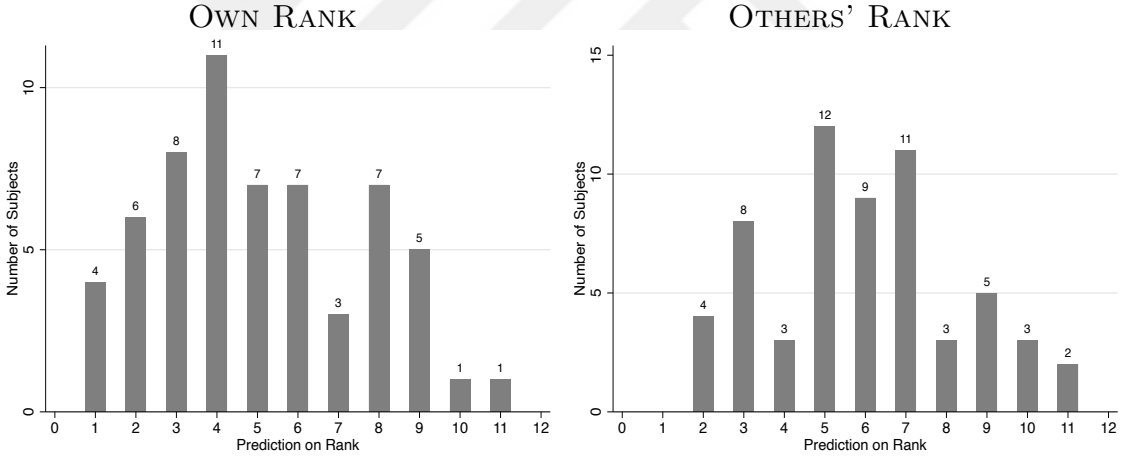
Notes: In Model 1 and 2, the dependent variable is the players' initial belief minus their actual rank. In Model 3 and 4, the dependent variable is the 1st player's prediction on their own rank upon receiving a private signal minus their actual rank. In Model 5 and 6, the dependent variable is the 2nd player's prediction on their own rank upon receiving a private signal minus their actual rank. In Model 7 and 8, the dependent variable is the 1st player's revised prediction on their own rank minus their actual rank.

Relevant versus Non-relevant Prediction

In this section, we assess whether the distribution of the predictions differ when the object of interest is self-relevant. We compare the predictions of the players on their own rank against the predictions made on other player's rank with the same level of knowledge[23] .

Figure 3.14 demonstrates the distribution of the predictions of the first player after receiving a private signal regarding the target player's rank. Accordingly, the first player predicts his own ranking on the left side and the other player's ranking on the right side of the Figure 3.14.

Figure 3.14: Player I's Prediction on Rank



Subjects assess their ranking higher with mean 5.07 ($sd = 2.55$) compared to the mean 5.93 ($sd = 2.35$) of other's rank (Mann-Whitney test, $p = 0.053$).

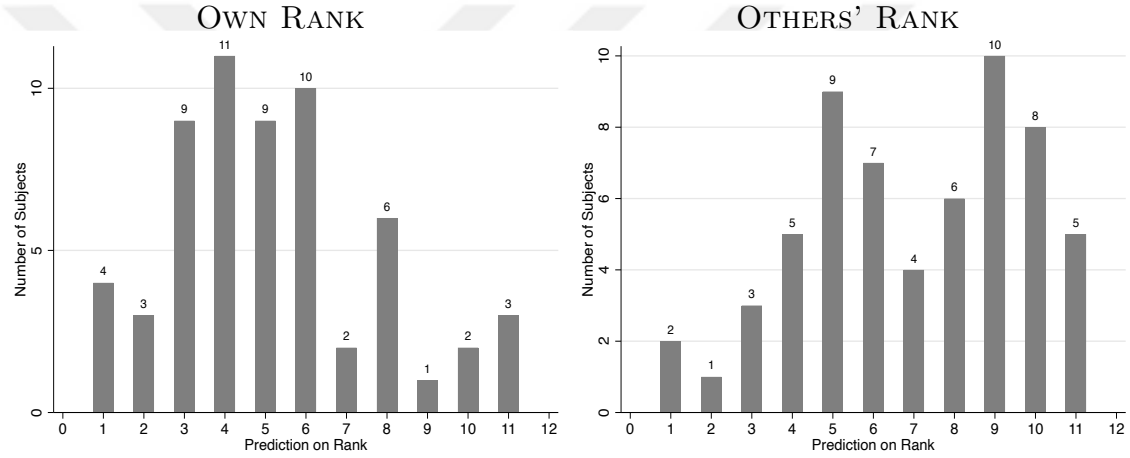
We analyze the extent to which 1st player's prediction deviates from the actual rank of the target subject, and discover that on average the prediction of the self-relevant item is more biased with mean value of -1.43 ($sd = 2.19$) than that of the

[23]Same level of knowledge refers to the case that if the player gets a private signal to predict his own rank, he will get only a private signal to predict other player's rank and vice versa. A detailed comparison of the 1st and 2nd players' predictions about the same target player's rank can be found in the section C.3

non-relevant item with mean value of -0.57 ($sd = 2.65$) on average (Mann-Whitney test, $p = 0.022$).

Figure 3.15 demonstrates the distribution of the predictions of the second player after receiving a private signal regarding the target player's rank and an information on the first player's prediction. The second player predicts her own ranking on the left side and the other player's ranking on the right side of the Figure 3.15.

Figure 3.15: Player II's Prediction on Rank



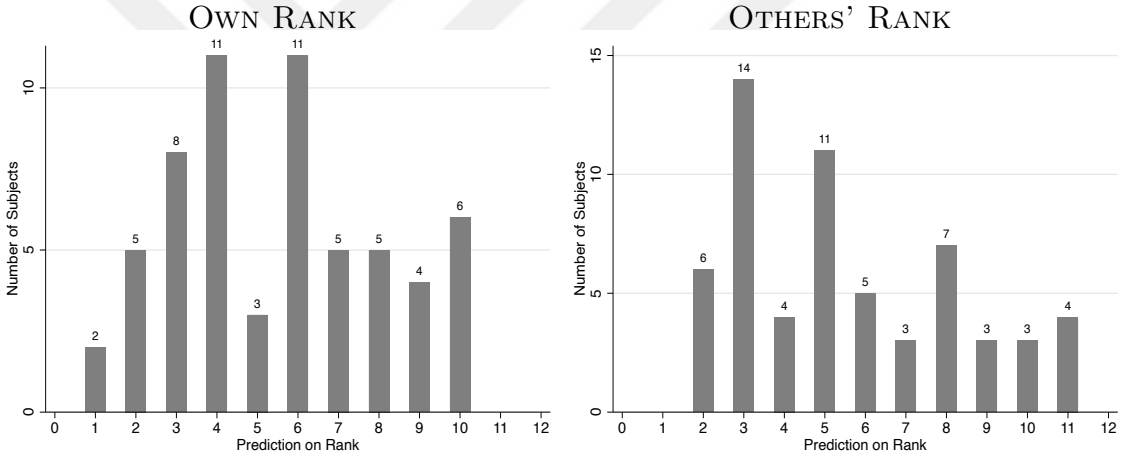
Subjects predict their ranking substantially higher with mean 5.17 ($sd = 2.55$) compared to the mean 7.02 ($sd = 2.70$) of other's rank (Mann-Whitney test, $p = 0.000$). While players assess their ranking marginally higher with mean 5.23 ($sd = 2.27$) compared to the mean 7.43 ($sd = 2.51$) of other's rank under imperfect information (Mann-Whitney test, $p = 0.001$), they estimate their rank with mean 5.10 (2.83) compared to the mean 6.60 (2.86) of other's rank under perfect information (Mann-Whitney test, $p = 0.036$).

We measure the amount of the deviation with respect to the 2$^{\text{nd}}$ player's prediction from the actual rank of the target subject and find that the prediction of the self-relevant item is more biased with mean value of -1.33 ($sd = 2.35$) than that of the non-relevant item 0.52 ($sd = 2.39$) on average (Mann-Whitney test, $p = 0.000$). The 2$^{\text{nd}}$ player overestimates her own rank by 1.13 ($sd = 2.27$) whereas underesti-

mates the other player's rank by 0.20 ($sd = 2.06$) under the imperfect information (Mann-Whitney test, $p = 0.061$). On the other hand, the second player overestimates her own rank by 1.53 ($sd = 2.45$) and underestimates the other player's rank by 0.83 ($sd = 2.68$) under perfect information (Mann-Whitney test, $p = 0.001$).

Figure 3.16 demonstrates the distribution of the revised predictions of the first player after receiving an information on the second player's prediction. The first player revises his prediction about his own ranking on the left side and the other player's ranking on the right side of the Figure 3.16.

Figure 3.16: Player I's Revised Prediction on Rank



Subjects estimate their ranking with mean 5.53 ($sd = 2.58$) compared to the mean 5.55 ($sd = 2.74$) of other's rank (Mann-Whitney test, $p = 0.855$). In particular, player estimate their ranking lower with mean 5.90 ($sd = 2.52$) compared to the mean 5.10 ($sd = 2.63$) of other's rank (Mann-Whitney test, $p = 0.187$) under imperfect information whereas higher with mean 5.17 ($sd = 2.63$) compared to the mean 6.00 ($sd = 2.91$) of other's rank under perfect information (Mann-Whitney test, $p = 0.290$).

We calculate the amount to which 1st player's revised prediction deviates from the actual rank of the target subject, and realize that the prediction of the self-relevant item is more biased with average value of -0.97 ($sd = 1.99$) than that of the non-

relevant item with average deviation of -0.95 ($sd = 2.57$) on average ($p = 0.000$). The 1st player overestimates both her own rank and the other player's rank by 1.33 ($sd = 2.14$) and 1.27 ($sd = 2.70$) under the imperfect information, respectively ($p = 0.698$). Similarly, he overestimates his own rank by 0.60 ($sd = 1.79$) and the other player's rank by 0.63 ($sd = 2.43$) under perfect information ($p = 0.482$).

Next, we run a series of OLS regressions in Table 3.10 where the dependent variable is the deviation of the player's prediction from the target player's actual rank.

In Models 1-4, the first player makes a prediction after receiving a private signal about the rank of the target player. We see that the signal type has a significant effect on the deviation of the first player's prediction from the target player's rank (Model 1 and 3). In particular, we see that the first player that receives a high signal overestimates the target player's rank compared to a first player that receive a low signal[24]. This effect is significant controlling for other factors only when the player makes a prediction about his own rank (Model 2). The gender does not have any significance on the deviation.

In Models 5-8, the second player makes a prediction after receiving a private signal about target player's rank and information about the first player's prediction. The signal type has a significant effect on the dependent variable only when the second player makes a prediction about her own rank (Model 5). Particularly, the second player with high signal overestimates her own rank compared a second player with low signal. The significance of this effect is robust to controlling factors (Model 6). The first player's prediction and the information level do not have any significant effect on the dependent variable[25]. Females are more conservative with their assessment compared to men when the second players make a prediction about their own rank (Model 6).

[24]The impact of high signal is -1.579 points with respect to the prediction on own rank while it is -1.201 points with respect to the prediction about the other's rank ($p = 0.222$).

[25]Controlling for other factors, the first player's prediction has a significant effect on the dependent variable when the second player makes a prediction about her own rank. The second player relatively overestimates her rank with top half rank (1-6) assessment by the first player compared to the bottom half rank (7-12) prediction by the first player.

In Models 9-12, the first player makes a revised prediction after receiving an information about the prediction of the second player. We observe that the signal type has a significant impact on the deviation of the first player's prediction from the target player's rank (Model 9 and 11). The first player that receives a high signal overestimates the target player's rank compared to a first player that receive a low signal[26]. This effect remains significant controlling for other factors only when the first player makes a revised prediction about his own rank (Model 10). The second player's prediction and the information level have a significant effect on the dependent variable only when the first player makes a revised prediction about the other player's rank (Model 11). The first player makes a relatively more optimistic revised prediction about the rank of the second player when the second player predicts to rank in the top-half (between 1 and 6) compared to bottom-half (between 7 and 12), and there is an imperfect information in comparison with the perfect information setting. The significance of these variables is robust to controlling factors (Model 12). The gender does not have any significance on the deviation.

[26]The effect of high signal is -1.094 points with respect to the revised prediction on own rank while it is -1.061 points with respect to the revised prediction on other's rank ($p = 0.548$).

Table 3.10: Deviation from the Target Player's Rank

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Deviation	1st Player	1st Player	1st Player	1st Player	2nd Player	2nd Player	2nd Player	2nd Player	1st Player	1st Player	1st Player	1st Player
High Signal	-1.579**	-1.529*	-1.201**	-0.990	-1.761**	-1.640**	-2.118	-1.908	-1.094**	-1.061**	-1.464*	-1.290
	(0.483)	(0.599)	(0.425)	(0.687)	(0.394)	(0.490)	(1.064)	(1.016)	(0.320)	(0.360)	(0.558)	(0.769)
Target Rank	-0.577***	-0.605***	-0.687***	-0.655***	-0.696***	-0.763***	-0.681***	-0.634***	-0.621***	-0.604**	-0.847***	-0.825***
	(0.057)	(0.078)	(0.087)	(0.126)	(0.106)	(0.075)	(0.102)	(0.079)	(0.112)	(0.159)	(0.069)	(0.079)
1st Player's Prediction					-0.829	-1.206**	-0.516	-0.639				
					(0.462)	(0.340)	(0.558)	(0.643)				
2nd Player's Prediction									-1.293	-1.226	-3.389***	-3.693**
									(0.739)	(0.901)	(0.572)	(1.226)
Information Level					0.0273	-0.330	0.250	0.536	0.171	0.188	0.957**	0.830*
					(0.399)	(0.438)	(0.658)	(0.649)	(0.200)	(0.142)	(0.274)	(0.323)
Female		0.486		-0.233		1.037**		0.021		0.472		-0.076
		(0.375)		(0.649)		(0.287)		(0.375)		(0.437)		(0.464)
Constant	3.108***	1.960	4.457***	3.004	4.496**	4.658	6.108***	2.213	4.112**	3.610	7.356***	7.780**
	(0.546)	(1.322)	(0.808)	(3.091)	(1.012)	(3.240)	(1.195)	(2.146)	(1.092)	(2.924)	(0.887)	(2.712)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	60	60	60	59	60	59	60	60	60	60	60	59
R^2	0.527	0.520	0.555	0.499	0.511	0.545	0.471	0.484	0.534	0.514	0.695	0.670

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

In Model 1 and 2, the dependent variable is the 1st player's prediction on their own rank upon receiving a private signal minus their own rank. On the other hand, in Model 3 and 4, the dependent variable is the 1st player's prediction about the 2nd player's rank upon receiving a private signal minus the 2nd player's rank. In Model 5 and 6, the dependent variable is the 2nd player's prediction on their own rank upon receiving a private signal minus their own rank. Instead, in Model 7 and 8, the dependent variable is the 2nd player's prediction about the 1st player's rank minus the 1st player's actual rank. In Model 9 and 10, the dependent variable is the 1st player's revised prediction on their own rank minus their own rank. In Model 11 and 12, the dependent variable is the 1st player's revised prediction about the 2nd player's rank minus the 2nd player's rank.

Table 3.11 shows the direction of the change in the predictions of the subjects compared to their previous prediction on their rank.

Table 3.11: Change in the Next Predictions

Breakdown	Group	PI's Prediction			PI's Revised Prediction			PII's Prediction		
		Increase	No Change	Decrease	Increase	No Change	Decrease	Increase	No Change	Decrease
Actual Rank	Top Half	13	12	5	5	15	10	10	8	12
	Bottom Half	3	11	16	2	12	16	4	3	23
Gender	Female	7	9	8	4	7	13	11	3	25
	Male	9	14	13	3	20	13	3	8	9
	Total	16	23	21	7	27	26	14	11	35

Notes: The table reports the frequency of the changes in the subsequent prediction by subjects according to the actual rank of the subject and the gender.

While the subjects better their prediction on their rank in 21% (37 out of 180) of all cases, their prediction remain same or becomes worse compared to their previous prediction for the remaining 79% (143 out of 180) cases.

The top half ranked subjects improve their prediction in a positive way for 31% (28 out of 90) of all predictions while the bottom half ranked subjects do so for 10% (9 out of 90) of all available cases (test of proportions, $p = 0.001$). On the other hand, the top half ranked subjects worsen their prediction for 30% (27 out of 90) of all cases whereas the bottom half ranked subjects do so only for 61% (55 out of 90) of all cases (test of proportions, $p = 0.000$).

With respect to the frequency of the genders, female subjects increase their prediction in 25% (22 out of 87) of all predictions while males do so in 16% (15 out of 92) of all available cases (test of proportions, $p = 0.138$). On the other hand, females decrease their prediction in 53% (46 out of 87) of all cases whereas males subjects do so only in 38% (35 out of 92) of all possible cases (test of proportions, $p = 0.046$).

We also run a series of OLS regressions in Table 3.12 where the dependent variable is the subtraction of the player's prediction from the player's previous prediction on the target player's rank. The positive values of the dependent variable indicate that subjects worsen their prediction about target player's rank whereas negative

values show that subjects better their prediction about the target player's rank compared to their previous prediction.

First, the player's previous prediction has a significant effect on the update extent of the player's next prediction (except in Model 7 and 8). The signal type has a significant effect on the extent of the update (Model 1 and 5). In particular, we see that player's that receive high signal (i.e. your rank is better than that of randomly chosen participant with rank 9) updates his prediction about his own rank in a positive way compared to a player that receive low signal. This effect is robust to additional controls (Model 2 and 6). The other player's prediction has a significant effect on the extent of the update when the subject updates his prediction on the target player's rank[27] (except in Model 5)[28]. This effect is significant controlling for other factors. Lastly, the information level and the gender do not have any significance on the degree of the update.

Result 8 *The players rank themselves higher compared to the other players when they are given a private signal. The difference between predicting your own rank and other player's rank vanishes when the 1st player is given a chance to revise his prediction.*

[27]The effect of the 2$^{\text{nd}}$ player's prediction being in the top-half rank is -1.352 points when the 1$^{\text{st}}$ player makes a revised prediction concerning his own rank while it is -0.943 points when he makes a revised prediction about the 2$^{\text{nd}}$ player's rank ($p = 0.310$).

[28]In Model 5, the second player updates her prediction about her own rank. Even though the first player's prediction does not have any significance on the magnitude of the update, it becomes significant controlling for other factors in Model 6.

Table 3.12: Determinants of the Update

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Update	1st Player	1st Player	1st Player	1st Player	2nd Player	2nd Player	1st Player	1st Player
Previous Prediction	-0.690**	-0.707***	-0.243**	-0.211**	-0.517***	-0.584***	-0.0950	-0.0697
	(0.162)	(0.152)	(0.0590)	(0.0566)	(0.0948)	(0.109)	(0.0642)	(0.0606)
High Signal	-2.950***	-2.636**			-2.377***	-2.229**		
	(0.555)	(0.713)			(0.308)	(0.538)		
1st Player's Prediction					-1.138	-1.332**		
					(0.590)	(0.383)		
2nd Player's Prediction			-1.352**	-1.317**			-0.943*	-1.404**
			(0.296)	(0.343)			(0.437)	(0.361)
Information Level			0.058	0.076	0.319	0.022	-0.005	-0.233
			(0.271)	(0.360)	(0.239)	(0.267)	(0.500)	(0.583)
Female		-0.173		0.005		0.728		0.210
		(0.243)		(0.186)		(0.541)		(0.417)
Constant	5.096***	4.483	2.279***	2.813	4.610***	3.000*	0.906	8.696
	(1.047)	(4.403)	(0.443)	(2.313)	(0.701)	(1.356)	(0.718)	(5.552)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	60	60	60	60	60	59	60	59
R^2	0.439	0.465	0.232	0.197	0.515	0.528	0.017	0.135

OLS Regressions, standard error in parantheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: In Model 1 and 2, the dependent variable is the 1st player's prediction on their own rank upon receiving a private signal minus their initial belief on their rank. In Model 3 and 4, the dependent variable is the 1st player's revised prediction on their own rank upon receiving a private signal minus their previous prediction on their rank. In Model 5 and 6, the dependent variable is the 2nd player's prediction on their own rank upon receiving a private signal minus their initial belief on their rank. In Model 7 and 8, the dependent variable is the 1st player's revised prediction about 2nd player's rank minus their previous prediction about 2nd player's rank.

3.4 Discussion and Concluding remarks

Individuals often acquire information about both their own qualities and those of others through real-life experiences, forming their posterior beliefs. It is crucial to discover the potential biases in these subsequent beliefs across self-relevant and self-irrelevant contexts, as these beliefs impact critical decisions within organizational environments. As an example, a biased performance feedback of an employee by the manager might influence the future performance of the respective employee. Motivated by this, we explore the response to the relative performance feedback across self-relevant and self-irrelevant contexts in a social network setting. To this end, we conduct a social network experiment in Istanbul and do not find enough evidence to indicate that the structure of the network has any effect on the results when the predicted item is not self-relevant to any players in the network. We also find that the predictions becomes biased when the predicted item is self-relevant to at least one party in the network.

One limitation of this study is that we do not randomize the order of the tasks in the experiment. This in turn might have generated an order effect with respect to the results of the first two stage of experiment (section 3.2.1 and section 3.2.2).

We ask subjects to solve a series of questions in Raven's Progressive Matrices test and then predict their own rank in the test. A potential issue with the design could be such that the payoffs from Raven's Progressive Matrices (section 3.2.3) and beliefs (predictions) with respect to the ranking from Circular Social Learning on Rank (section 3.2.4) might possibly give rise to moral hazard. If subjects do not answer questions in the test intentionally, then they may hold a belief to rank last in the test to benefit maximum payoffs from the predictions. Upon inspection, we notice that subjects at least solve 13 (57%) questions correctly whereas no subject rank itself at the bottom place as 12^{th}; therefore, we conclude that the data do not suggest any evidence on the presence of moral hazard[29].

[29]Only 4 (3.3%) subjects hold prior beliefs to be in the bottom quarter rank, between 10^{th}-12^{th} places. Furthermore, after receiving a private signal only 2 (1.7%) subjects predict to be at the bottom quarter rank. Finally, no subjects rank itself to be at the bottom quarter rank with respect to their revised prediction.

Compensation for rank predictions is carried out using the quadratic scoring rule, known for its incentive compatibility [Selten, 1998]. However, there are some concerns regarding the incentive compatibility of this scoring rule as it is not invariant to subjects' risk preferences [Antoniou et al., 2015]. Although this elicitation mechanism aligns incentives solely for risk-neutral agents, we believe that participants convey their genuine beliefs. Firstly, subjects are presented with the precise formula and told that their optimal average earnings result from honestly reporting their true beliefs. Secondly, the payment range spans from 29 to 150 tokens, a spectrum where subjects are likely to exhibit risk-neutrality[30].

We elicit subjects initial belief on their own rank in the test and notice that the subjects overestimate their rank, a widely known phenomenon regarded as overconfidence in the literature [Burks et al., 2013; Skala, 2008]. Moreover, we observe that the bottom half ranked subjects are more biased with their initial assessment compared to the top half ranked subjects. We also find that women are less confident than men with their initial assessment notwithstanding both genders overestimate their rank.

After the elicitation of initial beliefs, subjects receive information in the form of either private signal or other player's prediction, or possibly both to make their predictions. As subjects receive a new piece of information, they become less and less biased with their subsequent predictions compared to their initial assessment on their rank. Although these later predictions remain an overestimation of the subjects' actual rank, the degree of the overestimation decreases with each new piece of information. Moreover, top-half ranked subjects overestimate their rank significantly less compared to the bottom-half ranked subjects though the part of the difference might be attributed to the ceiling effects for the top half ranked player.

Lastly, we compare the predictions of the subjects on their own rank against the predictions on the other participant's rank in order to see whether the self-relevance plays a role in the predictions. On average, the subjects rank themselves

[30]The formula for the payment of the rank predictions is as follows: $150 - (prediction - rank)^2$. Accordingly, the minimum payoff would be 29 tokens with perfectly inaccurate prediction and the maximum payoff would be 150 tokens with perfectly accurate prediction.

higher compared to the other participants when they are provided with the same set of information in the earlier predictions. Furthermore, the prediction of the self-relevant item is more biased than the non-relevant item. On the other hand, we find no difference between the prediction of one's own rank and the other's rank with respect to the revised predictions. Thus, we conclude that the difference between predicting one's own rank and other player's rank disappears when the player is asked to revise his prediction.

Overall, our results suggest that gaining insights into how beliefs evolve when exposed to new information could aid in devising strategies to reduce gaps between the individual beliefs and the relative performance. We believe that our results could be used to further the performance evaluation where an employee and their manager provide feedback about each other's performance within an organizational setting. Future research could focus on the variations of network between individuals, potentially resulting in more accurate predictions of the self-relevant and self-irrelevant relative performance.

BIBLIOGRAPHY

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Acemoglu, D. and Robinson, J. A. (2013). *Why nations fail: The origins of power, prosperity, and poverty*. Broadway Business.

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

Anderson, L. R. and Holt, C. A. (1997). Information cascades in the laboratory. *The American economic review*, pages 847–862.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Antoniou, C., Harrison, G. W., Lau, M. I., and Read, D. (2015). Subjective bayesian beliefs. *Journal of Risk and Uncertainty*, 50:35–54.

Bandura, A. and Walters, R. H. (1977). *Social learning theory*, volume 1. Englewood cliffs Prentice Hall.

Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292.

Bigoni, M., Bortolotti, S., Casari, M., and Gambetta, D. (2019). At the Root of the North–South Cooperation Gap in Italy: Preferences or Beliefs? *Economic Journal*, 129(619):1139–1152.

Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., and Pancotto, F. (2016).

Amoral familism, social capital, or trust? the behavioural foundations of the italian north–south divide. *Economic Journal*, 126(594):1318–1341.

Bisin, A. and Verdier, T. (2017). On the joint evolution of culture and institutions. Technical report, National Bureau of Economic Research.

Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. (2008). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states. *American Economic Review*, 98(1):294–310.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Boyd, R. and Richerson, P. J. (2009). Voting with your feet: Payoff biased migration and the evolution of group beneficial behavior. *Journal of Theoretical Biology*, 257(2):331–339.

Brekke, K. A., Hauge, K. E., Lind, J. T., and Nyborg, K. (2011). Playing with the good guys. a public good game with endogenous group formation. *Journal of Public Economics*, 95(9-10):1111–1118.

Buchan, N. R., Brewer, M. B., Grimalda, G., Wilson, R. K., Fatas, E., and Foddy, M. (2011). Global social identity and global cooperation. *Psychological science*, 22(6):821–828.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2013). Overconfidence and social signalling. *Review of Economic Studies*, 80(3):949–983.

Burlando, R. M. and Guala, F. (2005). Heterogeneous agents in public goods experiments. *Experimental Economics*, 8:35–54.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The quarterly journal of economics*, 129(3):1409–1447.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American economic review*, 89(1):306–318.

Cameron, L., Chaudhuri, A., Erkal, N., and Gangadharan, L. (2009). Propensities to engage in and punish corrupt behavior: Experimental evidence from australia, india, indonesia and singapore. *Journal of Public Economics*, 93(7-8):843–851.

Cárdenas, J.-C., Dreber, A., Von Essen, E., and Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in colombia and sweden. *Journal of Economic Behavior & Organization*, 83(1):11–23.

Centola, D. (2010). The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197.

Charness, G., Rustichini, A., and Van de Ven, J. (2011). Self-confidence and strategic deterrence. Technical report, Tinbergen Institute Discussion Paper.

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.

Chugunova, M., Luhan, W. J., and Nicklisch, A. (2020). When to leave carrots for sticks: On the evolution of sanctioning institutions in open communities. *Economics Letters*, page 109155.

Cobo-Reyes, R., Katz, G., and Meraglia, S. (2019). Endogenous sanctioning institutions and migration patterns: Experimental evidence. *Journal of Economic Behavior & Organization*, 158:575–606.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395.

Crawford, S. E. and Ostrom, E. (1995). A grammar of institutions. *American political science review*, pages 582–600.

Dannenberg, A. and Gallier, C. (2020). The choice of institutions to solve cooperation problems: a survey of experimental research. *Experimental Economics*, 23:716–749.

Dittmar, J. E. and Meisenzahl, R. R. (2020). Public Goods Institutions, Human Capital, and Growth: Evidence from German History. *The Review of Economic Studies*, 87(2):959–996.

Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current directions in psychological science*, 12(3):83–87.

Edwards, W. (1968). Conservatism in human information processing. *Formal representation of human judgment*.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.

Ertac, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, 133(4):1645–1692.

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

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868):137–140.

Fehr, E. and Williams, T. (2018). Social norms, endogenous sorting and the culture of cooperation.

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

Gächter, S. and Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):791–806.

Gächter, S., Herrmann, B., and Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2651–2661.

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science*, 322(5907):1510–1510.

Gächter, S. and Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531(7595):496.

Gächter, S. and Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, 3(2-3):303–314.

Gneezy, U., Gravert, C., Saccardo, S., and Tausch, F. (2017). A must lie situation–avoiding giving negative feedback. *Games and Economic Behavior*, 102:445–454.

Grimm, V. and Mengel, F. (2020). Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82.

Grossman, Z. and Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, 84(2):510–524.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2):23–48.

Gunnthorsdottir, A., Houser, D., and McCabe, K. (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization*, 62(2):304–315.

Gunnthorsdottir, A., Vragov, R., Seifert, S., and McCabe, K. (2010). Near-efficient equilibria in contribution-based competitive grouping. *Journal of Public Economics*, 94(11-12):987–994.

Gürdal, M. Y., Torul, O., and Yahşi, M. (2020). Heterogeneity in public good contributions: The pivotal role of the first round.

Gürerk, Ö. (2013). Social learning increases the acceptance and the efficiency of punishment institutions in social dilemmas. *Journal of Economic Psychology*, 34:229–239.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770):108–111.

Gürerk, Ö., Irlenbusch, B., and Rockenbach, B. (2014). On cooperation in open communities. *Journal of Public Economics*, 120:220–230.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53(1):3–35.

Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). "economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6):795–815.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., et al. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972):1480–1484.

Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868):1362–1367.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, 2nd edition.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

Huffman, D., Raymond, C., and Shvets, J. (2022). Persistent overconfidence and biased memory: Evidence from managers. *American Economic Review*, 112(10):3141–3175.

Isaac, R. M. and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199.

Kearns, M., Suri, S., and Montfort, N. (2006). An experimental study of the coloring problem on human subject networks. *science*, 313(5788):824–827.

Knauft, B. M. (1985). *Good company and violence: Sorcery and social action in a lowland New Guinea Society*. Number 3. Univ of California Press.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6):1121.

Lamba, S. and Mace, R. (2011). Demography and ecology drive variation in cooperation across human populations. *Proceedings of the National Academy of Sciences*, 108(35):14426–14430.

Möbius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence. *NBER Working paper*, 17014.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, 115(2):502.

Muthukrishna, M., Bell, A., Henrich, J., Curtin, C., Gedranovich, A., McInerney, J., and Thue, B. (2018). Beyond weird psychology: Measuring and mapping scales of cultural and psychological distance. ssrn.

Nicklisch, A., Grechenig, K., and Thöni, C. (2016). Information-sensitive leviathans. *Journal of Public Economics*, 144:1–13.

Ones, U. and Putterman, L. (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization*, 62(4):495–521.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2):404–417.

Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1993). Making democracy work: Civic institutions in modern italy.

Raven, J. et al. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment*, pages 223–237. Springer.

Richerson, P. J., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., et al. (2016). Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, 39.

Richerson, P. J. and Boyd, R. (2005). Not by genes alone : How culture transformed human evolution.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: An experimental study. *The American Economic Review*, pages 1068–1095.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1:43–61.

Skala, D. (2008). Overconfidence in psychology and finance-an interdisciplinary literature review. *Bank I kredyt*, (4):33–50.

Soltis, J., Boyd, R., and Richerson, P. J. (1995). Can group-functional behaviors evolve by cultural group selection?: An empirical test. *Current Anthropology*, 36(3):473–494.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta psychologica*, 47(2):143–148.

Thöni, C. (2019). Cross-cultural behavioral experiments: potential and challenges. In Schram, A. and Ule, A., editors, *Handbook of Research Methods and Applications in Experimental Economics*. Edward Elgar Publishing.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, 64(5):416–424.

Tuzin, D. (2013). *Social complexity in the making: A case study among the Arapesh of New Guinea*. Routledge.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

Weber, R. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review*, 96(1):114–126.

Weber, R. and Dawes, R. (2005). Behavioral economics. *The handbook of economic sociology*, pages 90–108.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1):110.

Zhang, N. (2018). Institutions, norms, and accountability: A corruption experiment with northern and southern italians. *Journal of Experimental Political Science*, 5(1):11–25.

Appendix A

APPENDIX: CULTURE AND PREVALENCE OF SANCTIONING INSTITUTIONS

Table A.1 demonstrates the main cultural and societal background of the societies of our subject pools.

Table A.1: Cultural and societal background of subject pools

		Social capital variables	Norms of civic cooperation					Secularity &Autonomy		Law enforcement & Democracy	
Subject Pool	Country	Share of people who can be trusted	Avoiding fare on transportation	Claiming government benefits	Stealing Property	Cheating on taxes	Accepting bribe	Overall Secular Values	Autonomy Index	Rule of law	Importance of democracy
Erfurt	Germany	0.45	9.08	9.06	9.68	-	9.31	0.41	0.54	1.68	8.94
Istanbul	Turkey	0.12	9.16	9.33	9.74	9.73	9.80	0.24	0.00	-0.03	8.57

Rule of law data is retrieved from World Bank between the years of 2010 and 2017 while the rest of data are acquired from World Value Survey Wave 6

Norms of civic cooperation score ranges between 1 and 10 (1=very weak norms of civic cooperation; 10=very strong norms of civic cooperation)

Secular value lies between 0 and 1 (1=very secular)

Autonomy index ranges between -1 to 1 (-1=weakest autonomy,1=highest autonomy)

Rule of law ranges approximately between -2.5 and 2.5 (2.5=very strong rule of law)

Importance of democracy ranges between 1 and 10 (1=not important at all, 10= absolutely important)

Table A.2 reflects the mean values and standard errors of contribution decisions and realized payoffs over the 1^{st} and 2^{nd} half of the game for the VCM and the PUN communities in Turkey and Germany. There is no significant difference in the average contributions in the first and second half of the game for the VCM and the PUN communities across countries.

Table A.2: Contributions and payoffs over the first and second half of the game

	Average Contribution				Average Payoff			
	VCM		PUN		VCM		PUN	
	TUR	GER	TUR	GER	TUR	GER	TUR	GER
Period 1	3.14	7.39	10.05	13.07	41.88	44.44	21.25	23.04
	(0.72)	(0.69)	(1.38)	(0.77)	(0.43)	(0.41)	(3.28)	(6.45)
1^{st} Half	2.46	3.12	16.88	17.59	41.48	41.87	39.57	40.26
	(0.39)	(0.53)	(0.85)	(0.63)	(0.23)	(0.32)	(2.50)	(2.13)
2^{nd} Half	1.21	2.00	19.54	19.57	40.73	41.20	49.14	48.84
	(0.83)	(0.84)	(0.24)	(0.19)	(0.50)	(0.51)	(0.86)	(0.88)
All Periods	2.23	2.92	18.48	18.83	41.34	41.75	45.33	45.64
	(0.52)	(0.57)	(0.43)	(0.31)	(0.31)	(0.34)	(1.23)	(1.13)

TUR and GER stand for the subjects in Turkey and Germany, respectively.

1^{st} Half includes Period 1-10 and Period 1-15 for Turkey and Germany, respectively.

2^{nd} Half contains Period 11-20 and Period 16-30 for Turkey and Germany, respectively.

The unit observation is a subject's own contribution to the public good for the mean statistic.

Standard errors are clustered across groups.

We define types of subjects according to the level of contribution in the first three period. In particular, we consider the level of contribution made in the PUN community for the first time in the first three period span to categorize the subjects. We label participants who are cooperators and also punish others as strong cooperators. Finally, we say that subjects who contribute 5 or less tokens are free-riders. Accordingly, Table A.3 indicates the mean level of contribution made by cooperators, strong cooperators and free-riders respectively. Averaged over all periods, there is no significant difference between free-riders, cooperators and strong cooperators across countries.

Table A.3: Contribution levels by different types of subjects

	Free-rider		Cooperator		Strong cooperator	
Statistic	TUR	GER	TUR	GER	TUR	GER
Mean	13.17	13.44	18.00	18.21	17.98	17.99
SE	(1.21)	(1.27)	(0.00)	(0.67)	(0.93)	(0.91)

TUR stands for the subjects in the respective category (e.g. free riders in Turkey)

GER stands for the subjects in such category (e.g. cooperators in Germany)

Figure A.1 reflects Hofstede [2001]'s six cultural dimensions for Turkey and Germany.

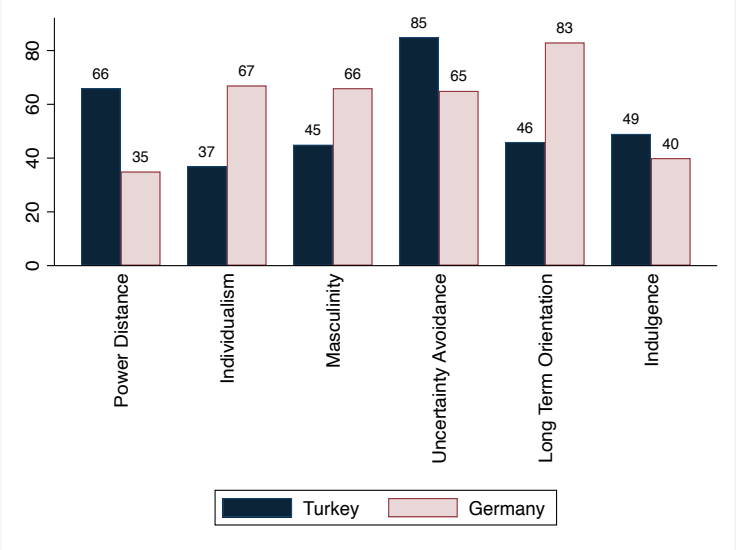


Figure A.1: Hofstede's Cultural Dimensions for Turkey and Germany

Figure A.2 shows the difference to the world mean in standard deviation of the six economic preferences for Germany and Turkey, retrieved from Falk et al. [2018].

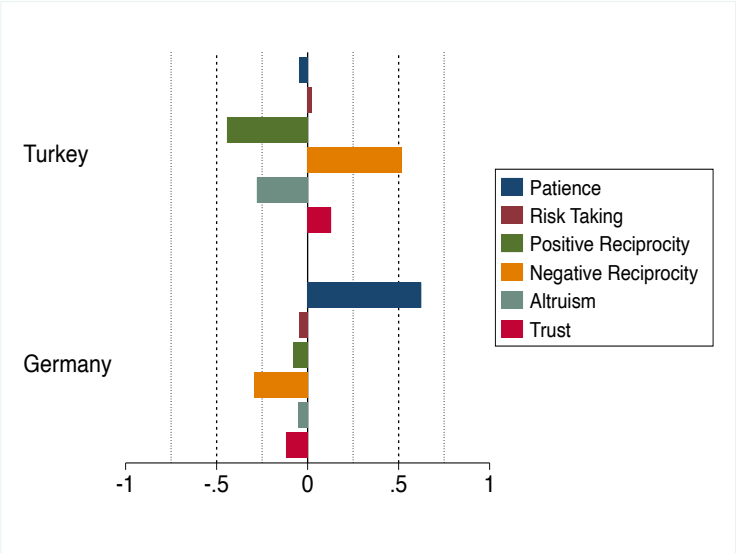
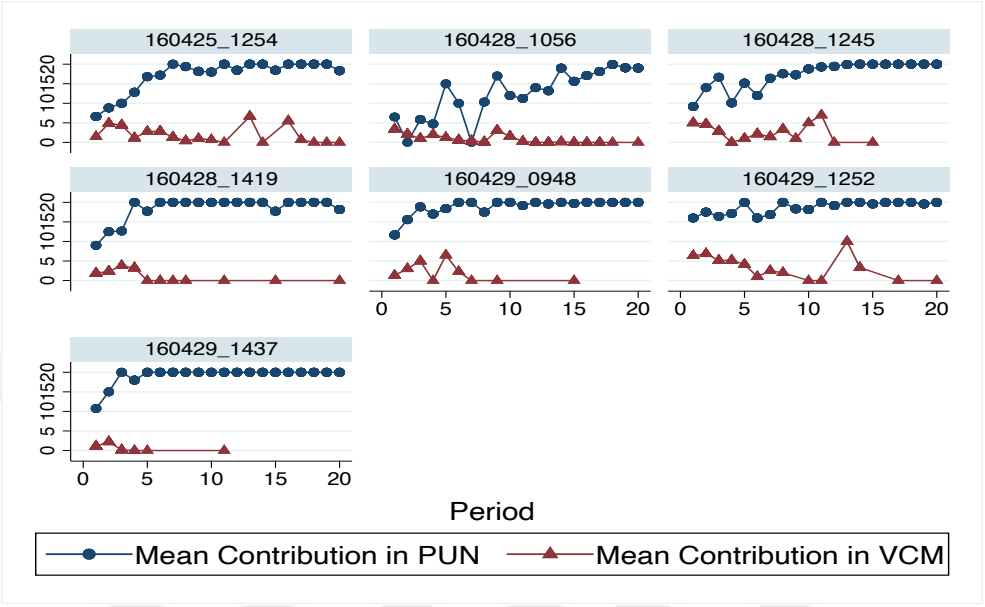
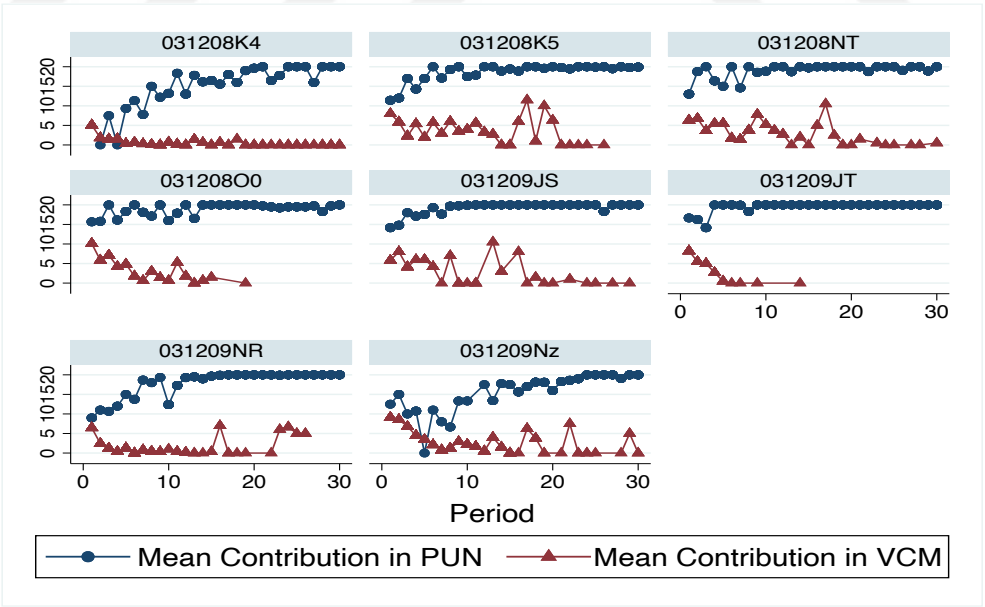


Figure A.2: Global Preferences Survey for Germany and Turkey

Figure A.3 illustrates the evolution of contributions per session in Turkey and Germany respectively.



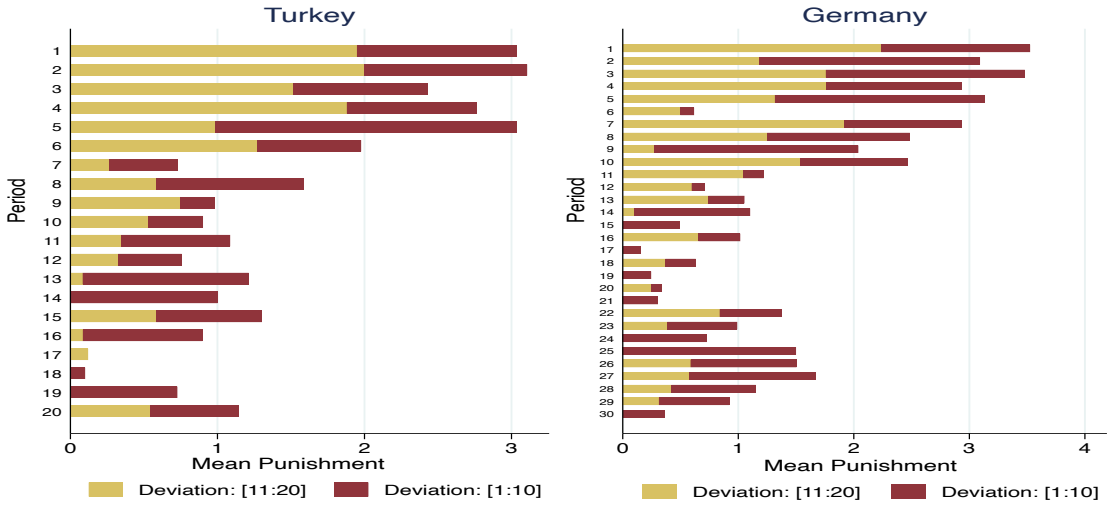
(a) Groups in Turkey



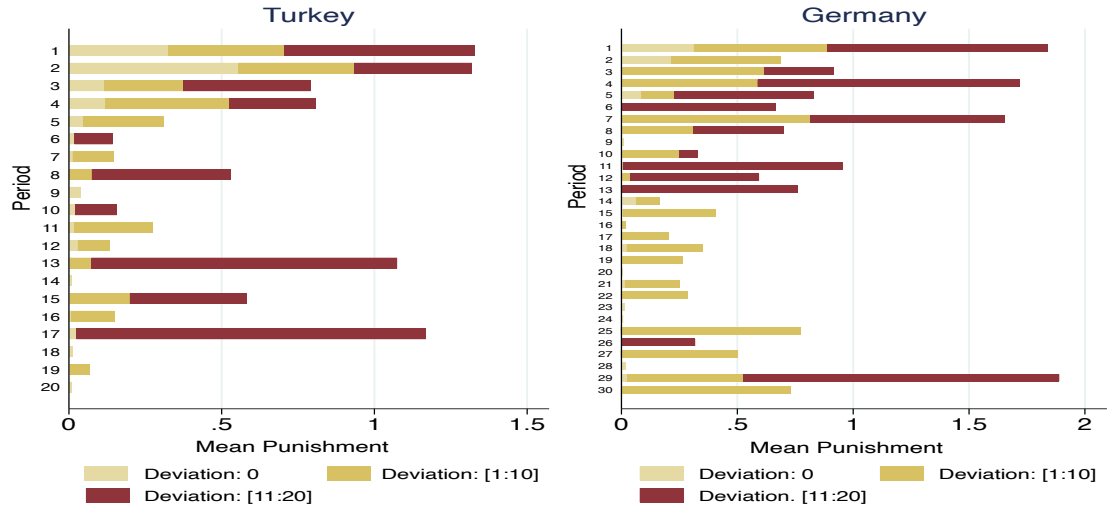
(b) Groups in Germany

Figure A.3: Evolution of contributions

Figure A.4 shows the mean punishment expenditures in Turkey and Germany per period. The upper portion of Figure A.4 indicates the average sanctioning points that fall into the category of free-rider punishment whereas the lower portion shows the anti-social punishment.



(a) Free-rider punishment



(b) Anti-social punishment

Figure A.4: Sanctioning Expenditures

Figure A.5 shows the average punishment tokens for each period in Turkey and Germany. The statistic denotes the mean punishment token with respect to all possible sanctioning decisions.

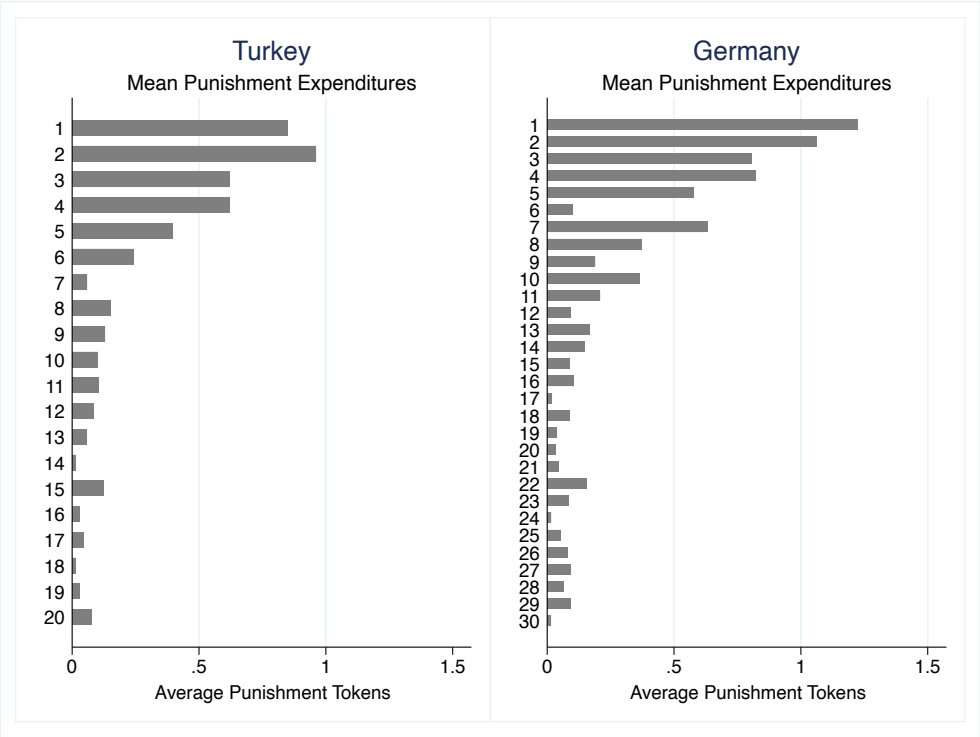


Figure A.5: Average punishment tokens

Figure A.6 shows the average frequency of anti-social punishment for each period in Turkey and Germany.

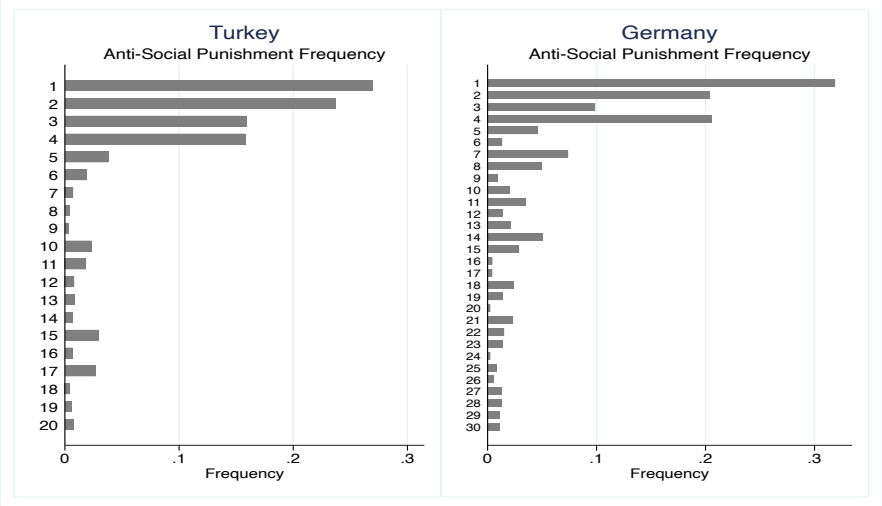


Figure A.6: Mean frequency of anti-social punishment

Figure A.7 shows the average anti-social punishment expenditures for each period in Turkey and Germany. The statistic denotes the mean value with respect to all possible anti social punishment decisions (e.g. anti-social punishment expenditure can be 0).

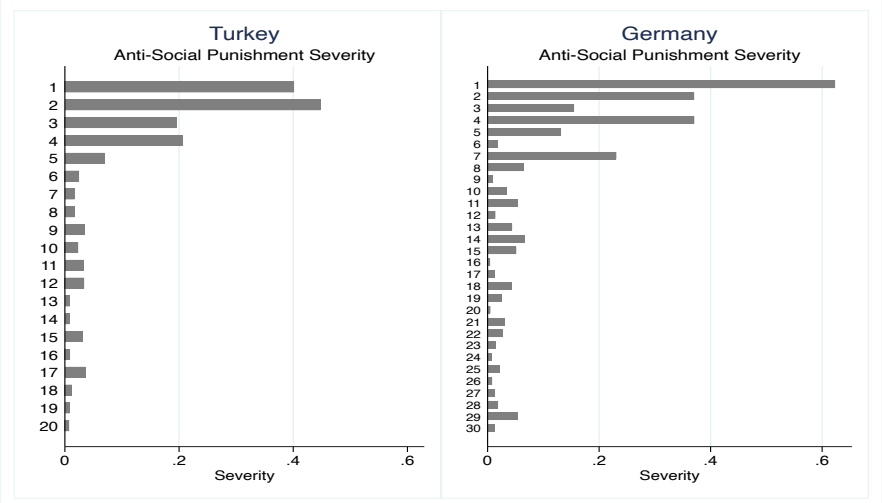


Figure A.7: Average anti-social punishment

Figure A.8 shows the average frequency of anti-social punishment for each period in Turkey and Germany by free-riders defined in Table 1.4.

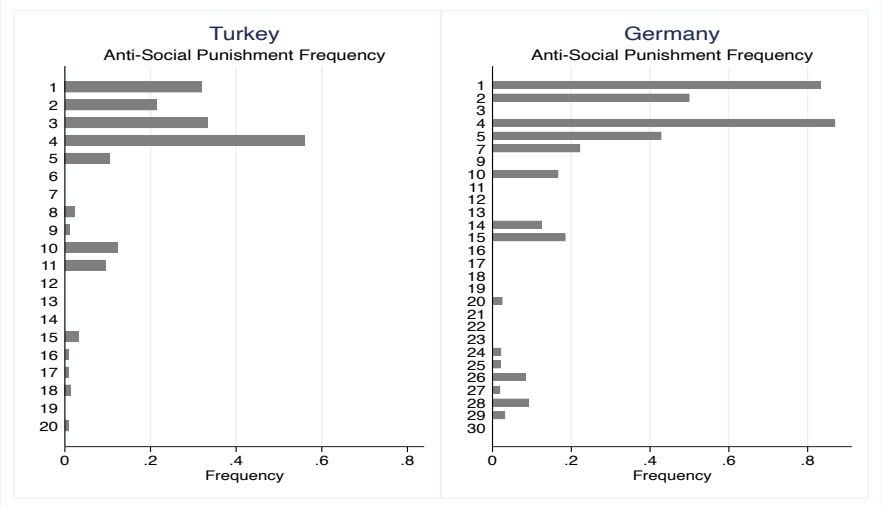


Figure A.8: Mean frequency of anti-social punishment for free-riders

Figure A.9 shows the average frequency of anti-social punishment for each period in Turkey and Germany by strong cooperators defined in Table 1.4.

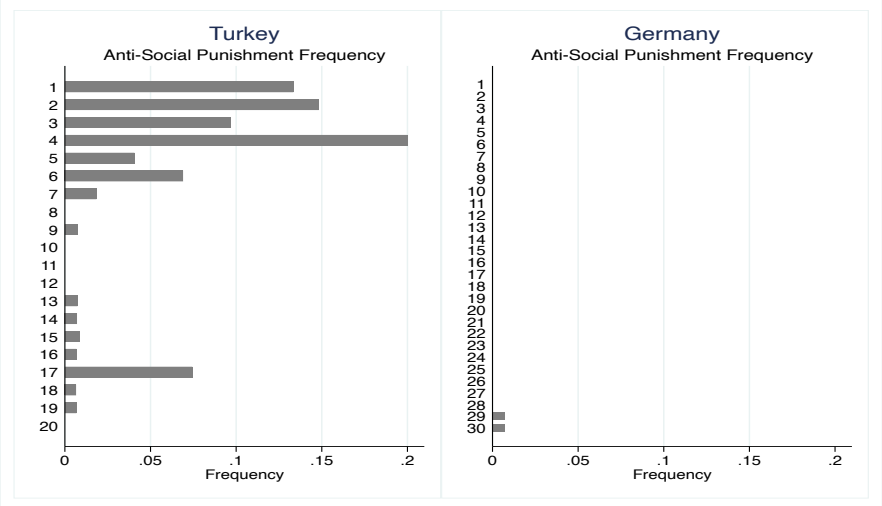


Figure A.9: Mean frequency of anti-social punishment for strong cooperators

Figure A.10 shows the average anti-social punishment expenditures for free-riders defined in Table 1.4 in each period in Turkey and Germany. The statistic denotes the mean value with respect to all possible anti social punishment decisions (e.g. anti-social punishment expenditure can be 0).

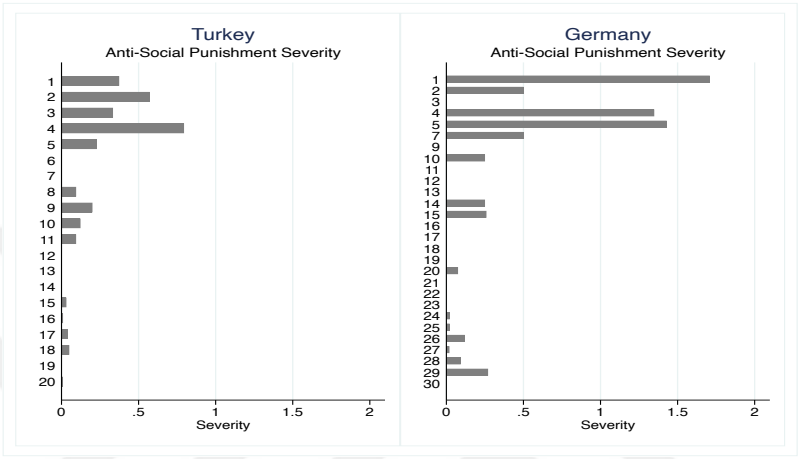


Figure A.10: Average anti-social punishment for free-riders

Figure A.11 shows the average anti-social punishment expenditures for strong cooperators defined in Table 1.4 in each period in Turkey and Germany. The statistic denotes the mean value with respect to all possible anti social punishment decisions (e.g. anti-social punishment expenditure can be 0).

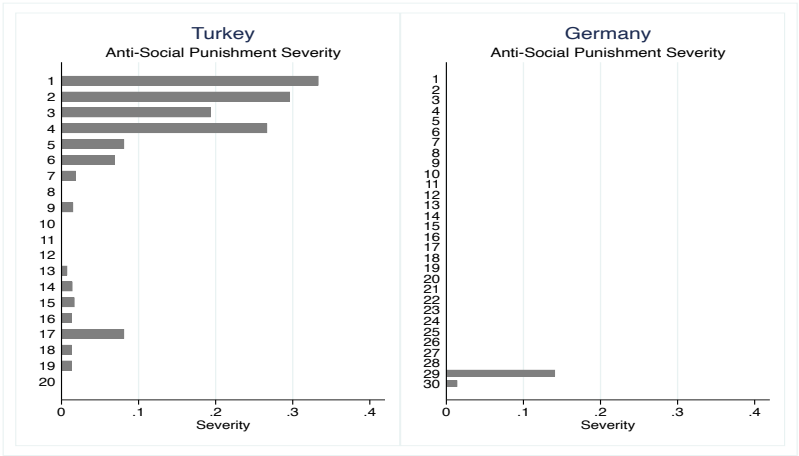


Figure A.11: Average anti-social punishment for strong cooperators

Figure A.12 shows the average frequency of anti-social punishment for each period in Turkey and Germany by free-riders defined in Table A.3.

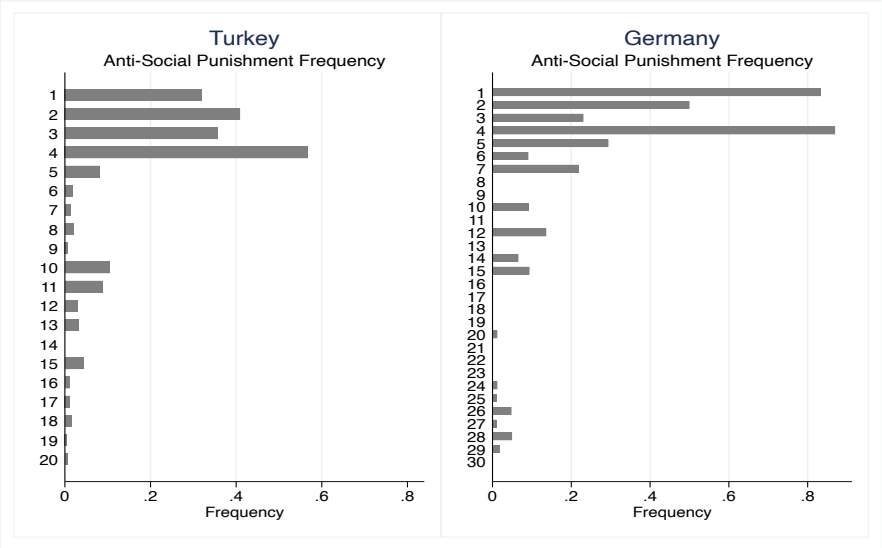


Figure A.12: Mean frequency of anti-social punishment for free-riders

Figure A.13 shows the average frequency of anti-social punishment for each period in Turkey and Germany by strong cooperators defined in Table A.3.

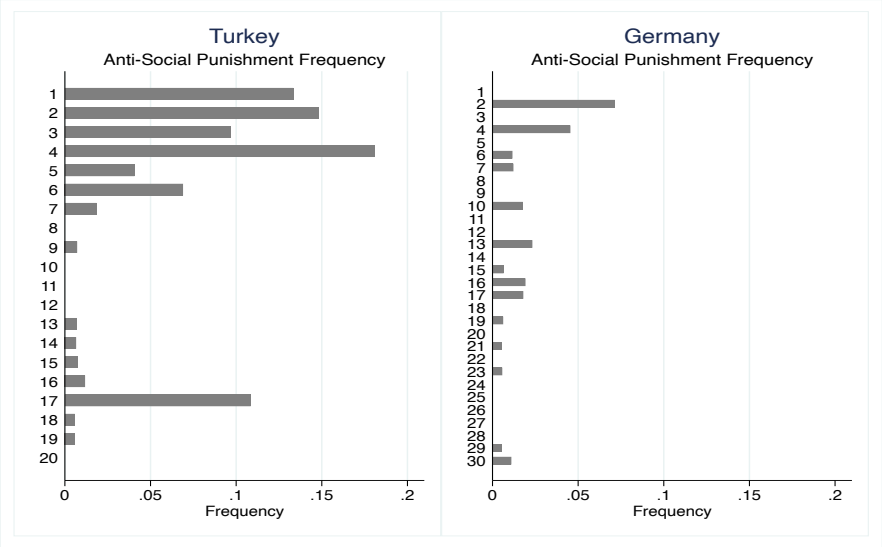


Figure A.13: Mean frequency of anti-social punishment for strong cooperators

Figure A.14 shows the average anti-social punishment expenditures for free-riders defined in Table A.3 in each period in Turkey and Germany. The statistic denotes the mean value with respect to all possible anti social punishment decisions (e.g. anti-social punishment expenditure can be 0).

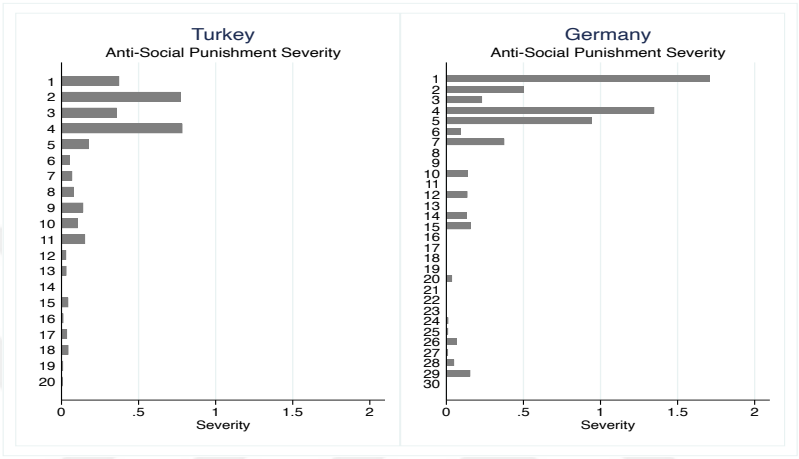


Figure A.14: Average anti-social punishment for free-riders

Figure A.15 shows the average anti-social punishment expenditures for strong cooperators defined in Table A.3 in each period in Turkey and Germany. The statistic denotes the mean value with respect to all possible anti social punishment decisions (e.g. anti-social punishment expenditure can be 0).

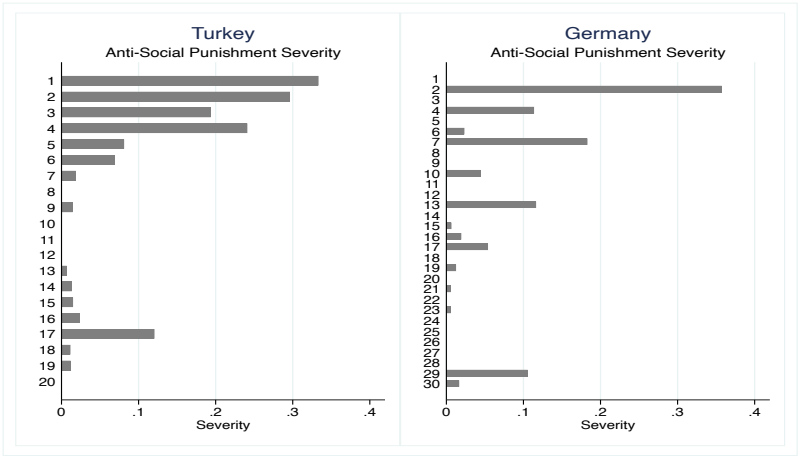
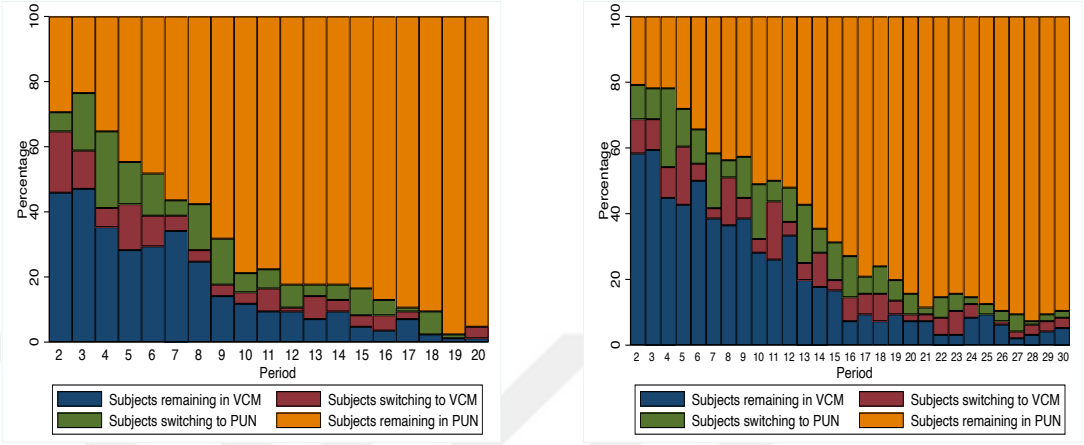


Figure A.15: Average anti-social punishment for strong cooperators

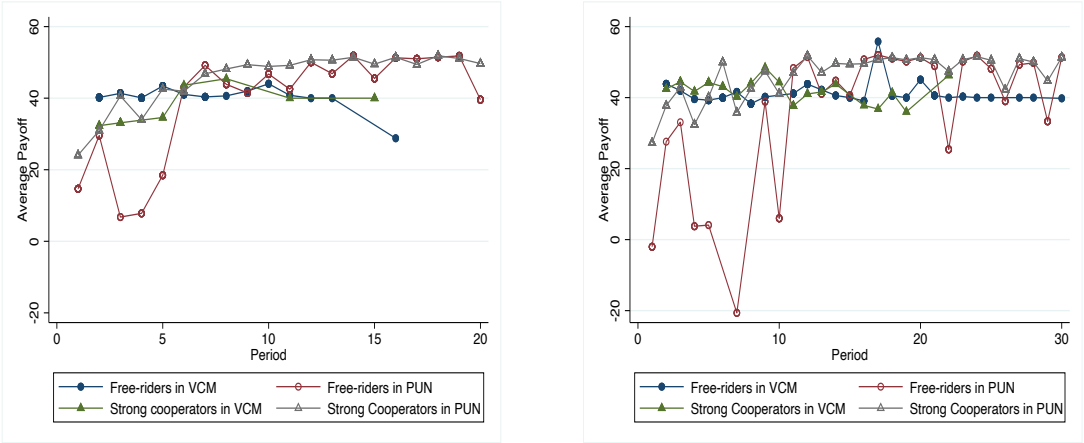
Figure A.16 shows the participants' community changing and remaining behavior for each period in Turkey and Germany.



(a) Turkey (b) Germany

Figure A.16: Evolution of community changing and remaining behavior

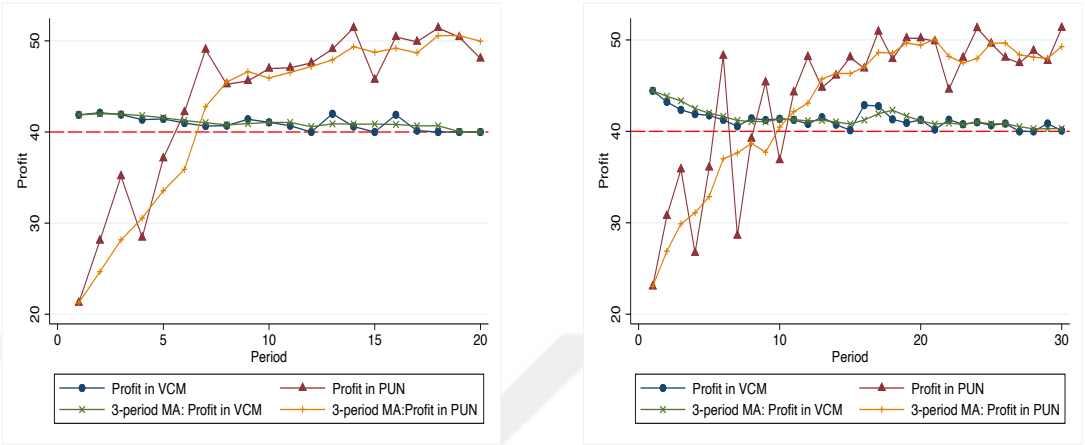
Figure A.17 shows the evolution of payoffs of participants, according to their types for each period in Turkey and Germany.



(a) Turkey (b) Germany

Figure A.17: Evolution of payoffs of different participant types

Figure A.18 shows the evolution of payoffs, i.e., period averages and moving averages (MA), from VCM and PUN communities.



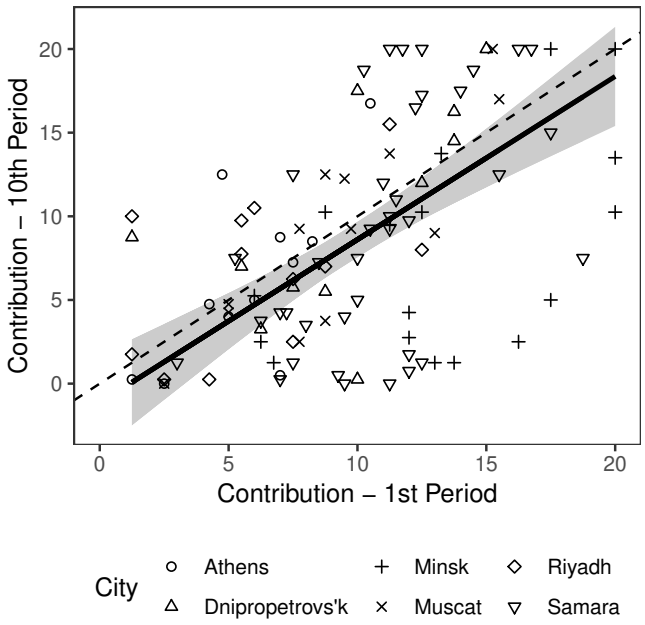
(a) Turkey (b) Germany

Figure A.18: Evolution of payoffs of different participant types

Appendix B

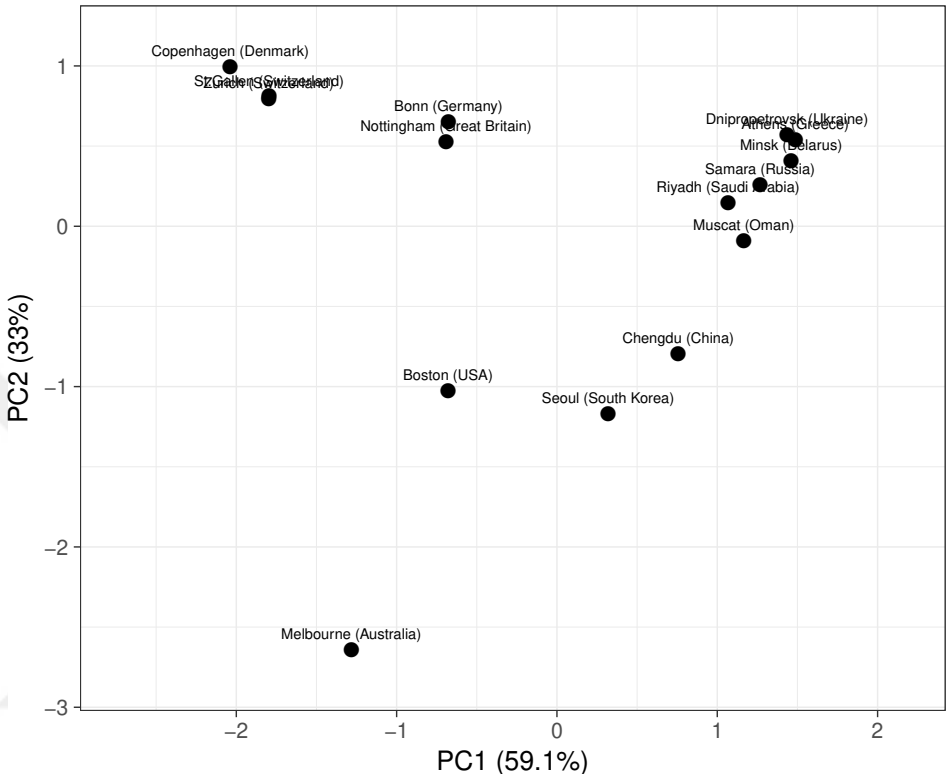
APPENDIX: THE SEEDS OF SUCCESS: THE PIVOTAL ROLE OF FIRST ROUND COOPERATION IN PUBLIC GOODS GAMES

Figure B.1: First-Period and Last-Period Contributions by City (FE) [Herrmann et al., 2008]



Notes: The horizontal axis denotes the 1st-period mean contribution, and the vertical axis denotes the 10th-period average contribution in 6 cities resembling Istanbul from the PCA analysis. The unit of observation is the average group contribution. The points below the dashed 45-degree line imply that those groups fail to improve upon the 1st period mean contribution while the points above the line show the groups that improve their mean contribution in the final period. The solid line denotes the linear fit, while the shaded gray areas display 95% confidence intervals.

Figure B.2: Principle Component Analysis [Herrmann et al., 2008]



Notes: Figure B.2 displays the principle component analysis scores of the fifteen cities by Herrmann et al., 2008 via 1) straight-line physical distance to Istanbul 2) GDP per capita (in 2017 current US Dollars) 3) cultural and psychological distance to Istanbul (Turkey) via Muthukrishna et al. [2018]'s WEIRD scale index.

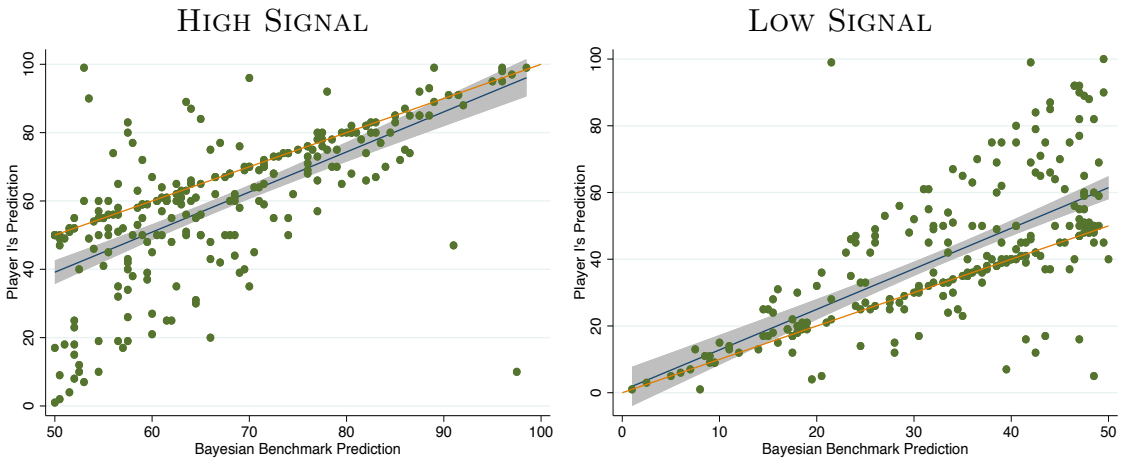
Appendix C

APPENDIX: SIGNALING YOURSELF

C.1 Players' Prediction by Signal Type

Figure C.1 demonstrates the predictions of the first player against the Bayesian benchmark predictions. The slope between the first player's prediction and the Bayesian benchmark prediction is 1.173 (with a standard error of 0.083) and 1.215 (with a standard error of 0.086) in the case of high and low signal, respectively ($p = 0.436$).

Figure C.1: Player I's Prediction by Signal Type

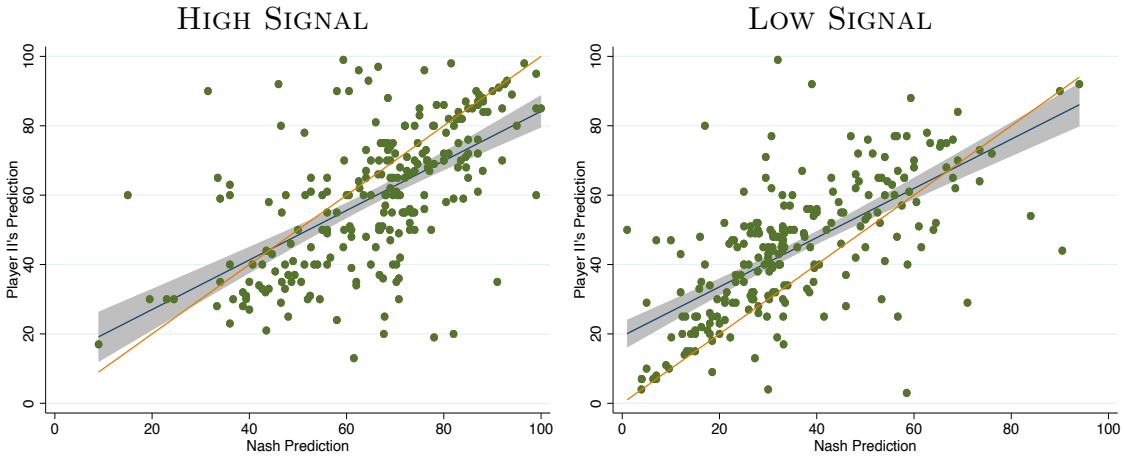


Notes: The horizontal axis denotes the prediction made by the 1st player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 1st player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

Figure C.2 shows the predictions of the second player against the Bayesian benchmark predictions. The slope between the second player's prediction and the Bayesian

benchmark prediction is 0.887 (with a standard error of 0.082) and 0.576 (with a standard error of 0.048) in the case of high and low signal, respectively ($p = 0.971$).

Figure C.2: Player II's Prediction by Signal Type

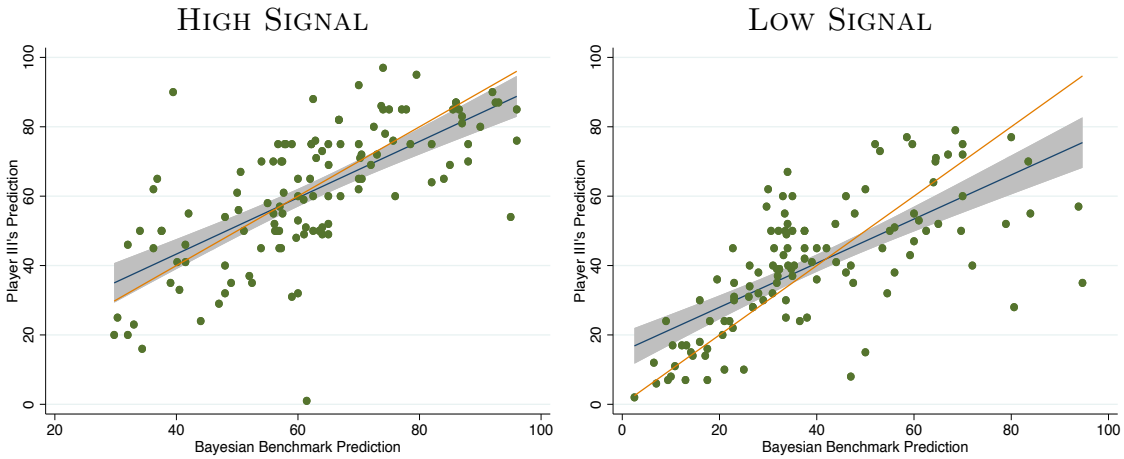


Notes: The horizontal axis stands for the prediction made by the 2nd player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 2nd player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

Figure C.3 demonstrates the predictions of the third player against the Bayesian benchmark predictions. The slope between the third player's prediction and the Bayesian benchmark prediction is 0.812 (with a standard error of 0.079) and 0.636 (with a standard error of 0.062) in the case of high and low signal, respectively ($p = 0.739$).

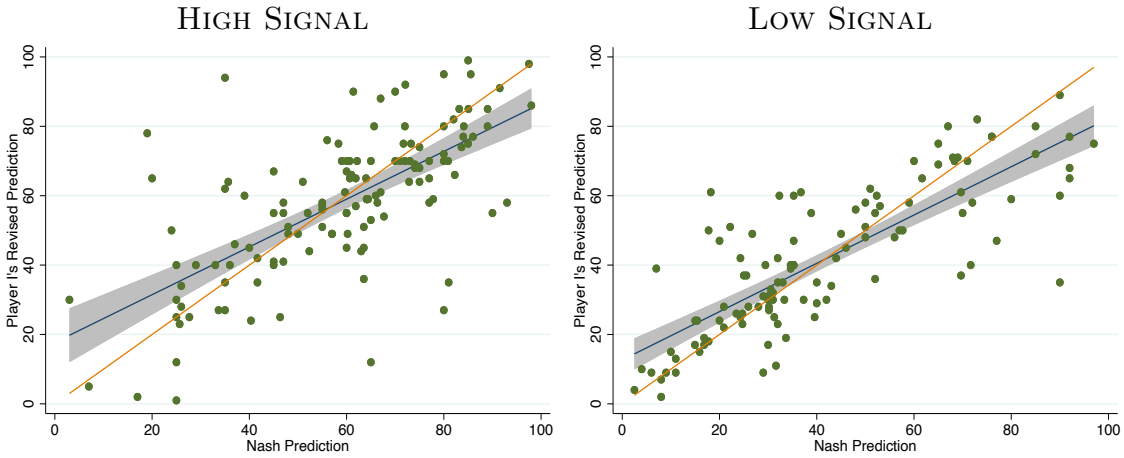
Figure C.4 shows the revised predictions of the first player against the Bayesian benchmark predictions. The slope between the first player's revised prediction and the Bayesian benchmark prediction is 0.688 (with a standard error of 0.066) and 0.696 (with a standard error of 0.049) in the case of high and low signal, respectively ($p = 0.971$).

Figure C.3: Player III's Prediction by Signal Type



Notes: The horizontal axis denotes the prediction made by the 3rd player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 3rd player's prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

Figure C.4: Player I's Revised Prediction by Signal Type



Notes: The horizontal axis stands for the revised prediction made by the 1st player and the vertical axis denotes the Bayesian benchmark Prediction. The orange line shows the 45 degree line where the Bayesian benchmark prediction coincides with the 1st player's revised prediction. The solid blue line denotes the linear fit while the shaded gray areas display 95% confidence intervals.

C.2 3rd Player - Imperfect Information

In this section, we analyze the Bayesian response of the third player in the linear social learning in the imperfect information scenario. Since 3rd player sees the prediction of the 1st and 2nd players partially; there can be 4 alternatives with respect to the imperfect information passed on to the 3rd player:

1. : $T_{p_1} > 50$, $T_{p_2} <= 50$

2. : $T_{p_1} <= 50$, $T_{p_2} > 50$

3. : $T_{p_1} > 50$, $T_{p_2} > 50$

4. : $T_{p_1} <= 50$, $T_{p_2} <= 50$

Accordingly, we find the Bayesian response of the 3rd player for each of the 4 alternatives below:

C.2.1 1st Case: $T_{p_1} > 50$, $T_{p_2} <= 50$

$$BR_{p_3}(s_3) = 75/2 \text{ if } T_{p_1} > 50 \text{ and } T_{p_2} <= 50 \text{ and } s_3 > t \text{ and } s_3 >= 75 \quad \text{(C.1)}$$

$$BR_{p_3}(s_3) = \frac{[((s_3/2) * (\frac{s_3}{(150-s_3)})) + ((2 * s_3/3) * ((225 - 3 * s_3)/(225 - 2 * s_3)))]}{[(s_3/(150 - s_3)) + ((225 - 3 * s_3)/(225 - 2 * s_3))]}$$

$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} <= 50 \text{ and } s_3 > t \text{ and } s_3 < 75 \quad \text{(C.2)}$$

$$BR_{p_3}(s_3) = \frac{((75 + s_3)/2)}{[(s_3/((75 + 2 * s_3)/3)) + ((75 - s_3)/(75 + s_3))]}$$

$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} <= 50 \text{ and } s_3 <= t \quad \text{(C.3)}$$

C.2.2 2nd Case: $T_{p_1} <= 50$, $T_{p_2} > 50$

$$BR_{p_3}(s_3) = 125/2$$

$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > 50 \text{ and } s_3 <= t \text{ and } s_3 <= 25 \quad \text{(C.4)}$$

$$BR_{p_3}(s_3) = \frac{[(((25+s_3)/2) * (\frac{(s_3-25)}{(175-s_3)}) + (((25+2*s_3)/3) * (\frac{(300-3*s_3)}{(275-2*s_3)})]}{[((s_3-25)/(175-s_3)) + ((300-3*s_3)/(275-2*s_3))]}$$

$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > 50 \text{ and } s_3 > t \quad \text{(C.5)}$$

$$BR_{p_3}(s_3) = \frac{[(((100+2*s_3)/3) * (\frac{(3*s_3-75)}{(25+2*s_3)}) + (((100+s_3)/2)) * (\frac{(100-s_3)}{(50+s_3)})]}{[((3*s_3-75)/(25+2*s_3)) + (\frac{(100-s_3)}{(50+s_3)})]}$$

$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} > 50 \text{ and } s_3 <= t \text{ and } s_3 > 25 \quad \text{(C.6)}$$

C.2.3 3rd Case: $T_{p_1} > 50$, $T_{p_2} > 50$

$$BR_{p_3}(s_3) = \frac{[(3*s_3/4)/(3*s_3/(3*s_3+100)) + (2*s_3/3) * (75/(2*s_3+75))]}{[(3*s_3/(3*s_3+100)) + (75/(2*s_3+75))]}$$

$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > 50 \text{ and } s_3 > t \text{ and } s_3 <= 75 \quad \text{(C.7)}$$

$$BR_{p_3}(s_3) = \frac{[((3*s_3/4) * (3*s_3/(3*s_3+100))) + (M_1) * (25/(M_1+25))]}{[(3*s_3/(3*s_3+100)) + (25/(M_1+25))]}$$

$$\text{if } T_{p_1} > 50 \text{ and } T_{p_2} > 50 \text{ and } s_3 > t \text{ and } s_3 > 75$$

$$\text{where } M_1 = \frac{[((75+s_3)/3) * ((s_3-75)/25) + (2*s_3/3) * ((100-s_3)/25)]}{[((s_3-75)/25) + ((100-s_3)/25)]} \quad \text{(C.8)}$$

$$BR_{p_3}(s_3) = \frac{[M_2 * (M_2/100) + M_3 * ((100 - M_3)/100)]}{[(M_2/100) + (100 - M_3)/100]}$$

if $T_{p_1} > 50$ and $T_{p_2} > 50$ and $s_3 < t$ and $s_3 > 75$

$$\text{where } M_2 = \frac{[\frac{(s_3 * s_3)}{((s_3+100)/2)} + \frac{(s_3*(200-2*s_3)}{(200+s_3))} + \frac{((300-3*s_3)*(300-s_3)}{(4*(300+s_3)))}]}{[\frac{(s_3*s_3)}{((s_3+100)*(s_3+100)/4)} + \frac{(s_3*(200-2*s_3))}{((200+s_3)*(200+s_3)/9)} + \frac{((300-3*s_3)*(300-3*s_3)}{((300+s_3)*(300+s_3)))}]}$$

$$\text{and } M_3 = \frac{[(s_3) + ((100 - s_3)/2)]}{[(3 * s_3/(100 + 2 * s_3)) + ((100 - s_3)/(100 + s_3))]} \quad \text{(C.9)}$$

$$BR_{p_3}(s_3) = \frac{[M_2 * (M_2/100) + M_4 * ((100 - M_4)/100)]}{[(M_2/100) + (100 - M_4)/100]}$$

if $T_{p_1} > 50$ and $T_{p_2} > 50$ and $s_3 < t$ and $75 >= s_3 >= 62.5$

$$\text{where } M_4 = \frac{[(s_3) + ((100 - s_3)/2)]}{[(3 * s_3/(100 + 2 * s_3)) + ((100 - s_3)/(100 + s_3))]} \quad \text{(C.10)}$$

$$BR_{p_3}(s_3) = \frac{[M_2 * (M_2/100) + M_5 * ((100 - M_5)/100)]}{[(M_2/100) + (100 - M_5)/100]}$$

if $T_{p_1} > 50$ and $T_{p_2} > 50$ and $s_3 < t$ and $62.5 >= s_3 >= 50$

$$\text{where } M_5 = \frac{[(50 * s_3/(125 - s_3)) + ((100 - s_3)/2)]}{[(100 * s_3/((75 + s_3) * (125 - s_3))) + ((100 - s_3)/(100 + s_3))]} \quad \text{(C.11)}$$

$$BR_{p_3}(s_3) = \frac{[M_2 * (M_2/100) + M_6 * ((100 - M_6)/100)]}{[(M_2/100) + (100 - M_6)/100]}$$

if $T_{p_1} > 50$ and $T_{p_2} > 50$ and $s_3 < t$ and $50 >= s_3$

$$\text{where } M_6 = \frac{[(50 * s_3/(125 - s_3)) + (150 - 2 * s_3) * 25/(150 - s_3)]}{[(\frac{(100 * s_3)}{((75+s_3)*(125-s_3))}) + (150 - 2 * s_3) * 75/((150 - s_3) * (150 + s_3))]} \tag{C.12}$$

C.2.4 4^{th} Case: $T_{p_1} <= 50$, $T_{p_2} <= 50$

$$BR_{p_3}(s_3) = \frac{[(((100 + 3 * s_3)/4) * \frac{(300-3*s_3)}{(400-3*s_3)}) + ((\frac{(100+2*s_3)}{3}) * (\frac{75}{(275-2*s_3)}))]}{[(300 - 3 * s_3)/(400 - 3 * s_3) + (75/(275 - 2 * s_3))]}$$

if $T_{p_1} <= 50$ and $T_{p_2} <= 50$ and $s_3 < t$ and $25 <= s_3$

(C.13)

$$BR_{p_3}(s_3) = \frac{[(((100 + 3 * s_3)/4) * \frac{(300-3*s_3)}{(400-3*s_3)}) + (N_1) * (25/(125 - N_1))]}{[(300 - 3 * s_3)/(400 - 3 * s_3) + (25/(125 - N_1))]}$$

if $T_{p_1} <= 50$ and $T_{p_2} <= 50$ and $s_3 <= t$ and $25 >= s_3$

$$\text{where } N_1 = \frac{[(((100 + 2 * s_3)/3) * ((s_3)/25)) + (((125 + s_3)/3) * ((25 - s_3)/25))]}{[((s_3)/25) + ((25 - s_3)/25)]}$$

(C.14)

$$BR_{p_3}(s_3) = \frac{[N_2 * ((100 - N_2)/100) + N_3 * N_3/100]}{[((100 - N_2)/100) + N_3/100]}$$

if $T_{p_1} <= 50$ and $T_{p_2} <= 50$ and $s_3 > t$ and $25 >= s_3$

$$\text{where } N_2 = \frac{[\frac{((s_3/4)*((3*s_3*3*s_3)}{((400-s_3)*(400-s_3))))} + \frac{((s_3/3)*((6*s_3*(100-s_3))}{((300-s_3)*(300-s_3))))} + \frac{((s_3/2)*((4*(100-s_3)*(100-s_3))}{((200-s_3)*(200-s_3))))}]}{[\frac{((3*s_3*3*s_3)}{((400-s_3)*(400-s_3)))} + \frac{((6*s_3*(100-s_3))}{((300-s_3)*(300-s_3)))} + \frac{((4*(100-s_3)*(100-s_3))}{((200-s_3)*(200-s_3)))}]}$$

$$\text{and } N_3 = \frac{[(s_3/2) * ((s_3)/((200 - s_3))) + (2 * s_3/3) * ((\frac{(300-3*s_3)}{(300-2*s_3)}))]}{[((s_3)/((200 - s_3))) + ((\frac{(300-3*s_3)}{(300-2*s_3)}))]}$$

(C.15)

$$BR_{p_3}(s_3) = \frac{[N_2 * ((100 - N_2)/100) + N_4 * N_4/100]}{[((100 - N_2)/100) + N_4/100]}$$

if $T_{p_1} <= 50$ and $T_{p_2} <= 50$ and $s_3 > t$ and $s_3 <= 75/2$ and $25 <= s_3$

$$\text{where } N_4 = \frac{[(s_3/2) * ((s_3)/((200 - s_3))) + (2 * s_3/3) * ((\frac{(300-3*s_3)}{(300-2*s_3)}))]}{[((s_3)/((200 - s_3))) + ((\frac{(300-3*s_3)}{(300-2*s_3)}))]}$$

(C.16)

$$BR_{p_3}(s_3) = \frac{[N_2 * ((100 - N_2)/100) + N_5 * N_5/100]}{[((100 - N_2)/100) + N_5/100]}$$

$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} <= 50 \text{ and } s_3 > t \text{ and } s_3 <= 50 \text{ and } 75/2 <= s_3$$

$$\text{where } N_5 = \frac{[(s_3/2) * ((s_3)/((200 - s_3))) + (200 - 2 * s_3) * 25/(175 - s_3)]}{[((s_3)/((200 - s_3))) + (200 - 2 * s_3) * 50/((175 - s_3) * (s_3 + 25))]} \tag{C.17}$$

$$BR_{p_3}(s_3) = \frac{[N_2 * ((100 - N_2)/100) + N_6 * N_6/100]}{[((100 - N_2)/100) + N_6/100]}$$

$$\text{if } T_{p_1} <= 50 \text{ and } T_{p_2} <= 50 \text{ and } s_3 > t \text{ and } 50 <= s_3$$

$$\text{where } N_6 = \frac{[(2 * s_3 - 50) * 25/(250 - s_3) + (200 - 2 * s_3) * 25/(175 - s_3)]}{[\frac{(2 * s_3 - 50) * 75}{((250 - s_3) * (s_3 + 50))} + (200 - 2 * s_3) * 50/((175 - s_3) * (s_3 + 25))]} \tag{C.18}$$

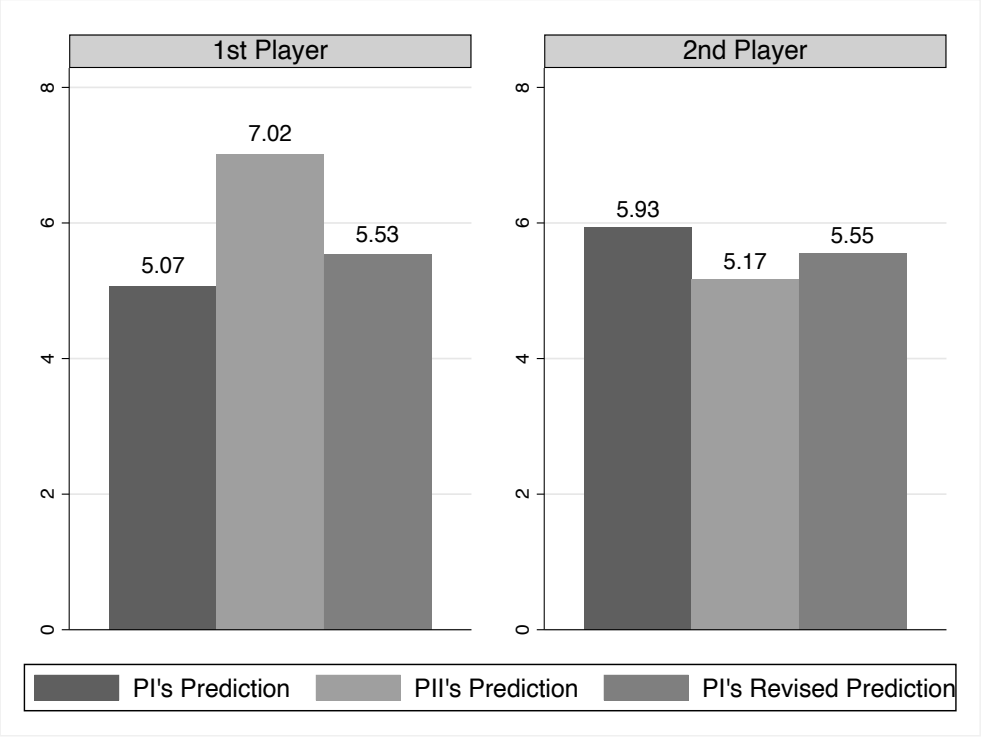
C.3 Comparison of the Predictions on Target Player's Rank

Figure C.5 shows the predictions on the target player's rank. In particular, players predict the 1st player's rank on the left side and the 2nd player's rank on the right side of the Figure C.5.

Focusing on the rank of the first player, we see that the second player's prediction is 1.95 worse than the first player's prediction (Mann-Whitney test, $p = 0.000$). The first player's revised prediction, however, is 1.49 better than the second player's prediction (Mann-Whitney test, $p = 0.003$).

When it comes to the actual rank of the second player, we observe that the second player's prediction is 0.76 better than the first player's prediction (Mann-Whitney

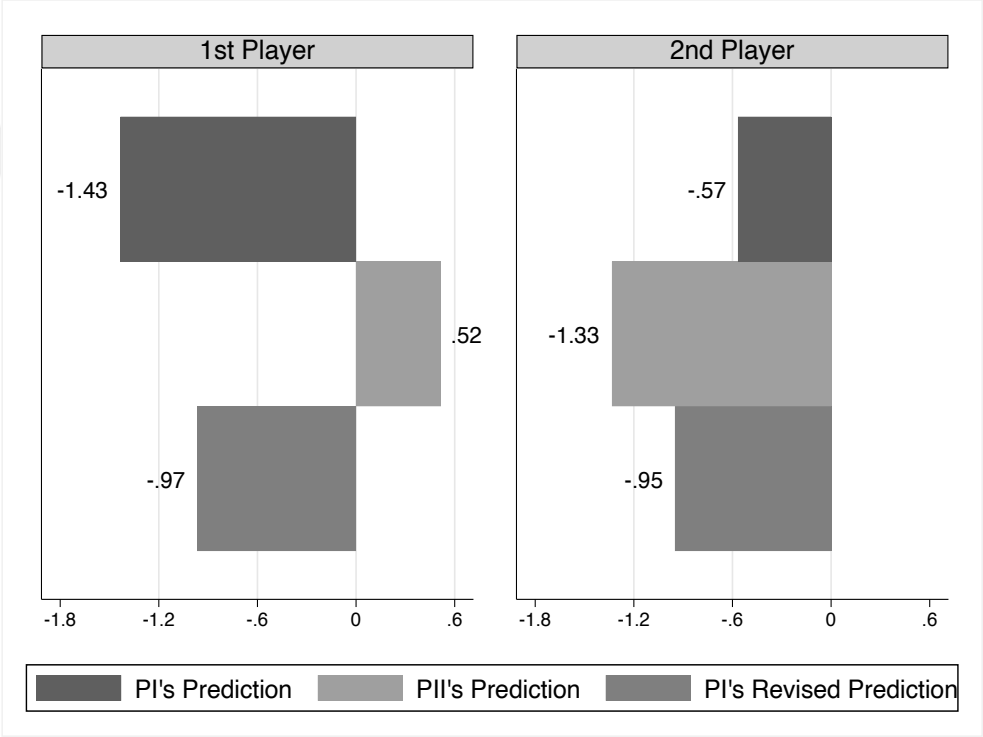
Figure C.5: Prediction on Target Player's Rank



test, $p = 0.062$). On the other hand, the revised prediction of the first player is 0.38 worse than the prediction of the second player and insignificant (Mann-Whitney test, $p = 0.617$).

Figure C.6 shows the deviation of the predictions on the target player's rank. In particular, the left side of the figure demonstrates the deviation of the predictions from the 1st player's rank whereas the right side illustrates the deviation of the predictions from the 2nd player's actual rank.

Figure C.6: Deviation on Target Player's Rank



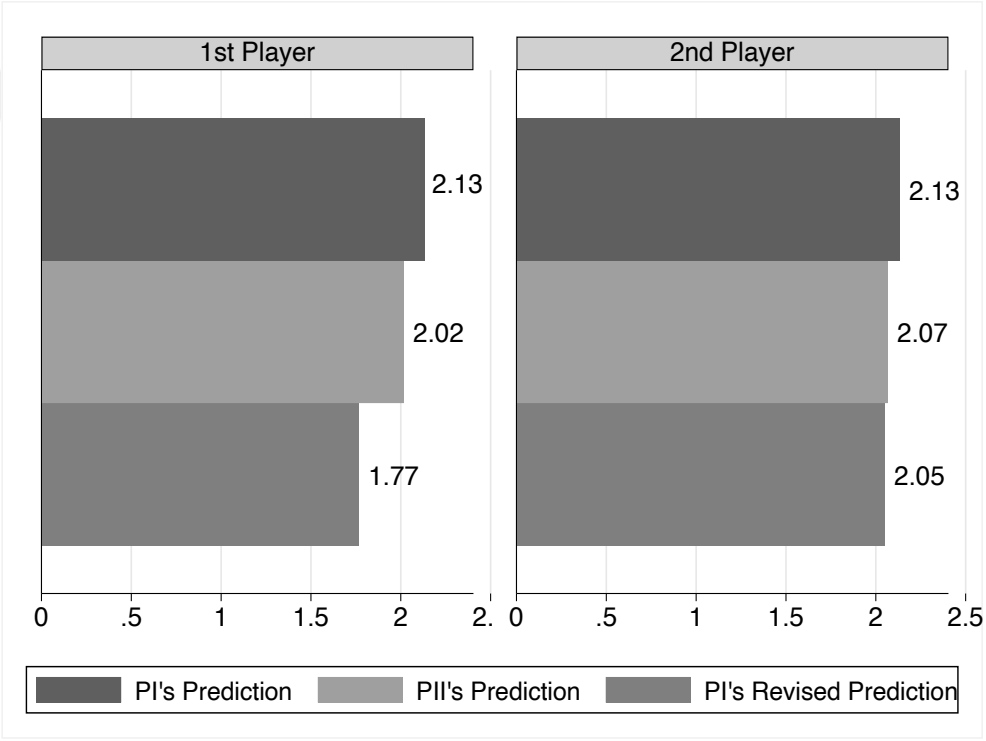
While the first player overestimates his own rank by 1.43, the second player underestimates it by 0.52 on average (Mann-Whitney test, $p = 0.000$). The first player also overestimates his own rank by 0.97 with his revised prediction, and the difference between this prediction and the second player's prediction is significant (Mann-Whitney test, $p = 0.001$).

For the prediction of the second player's rank, we notice that the second player overestimates her own rank by 1.33 whereas the first player overestimates it by 0.57 on average (Mann-Whitney test, $p = 0.064$). On the other hand, the first player overestimates the rank of the second player by 0.95 with respect to his revised prediction. The difference between the deviation from the 1st player's revised prediction

and the 2nd player's prediction is not significant (Mann-Whitney test, $p = 0.271$).

Figure C.7 shows the absolute deviation of the predictions from the target player's rank. In particular, the left side of the figure demonstrates the absolute deviation of the predictions from the 1st player's rank whereas the right side illustrates the absolute deviation of the predictions from the 2nd player's actual rank.

Figure C.7: Absolute Deviation on Target Player's Rank



The absolute deviation of the first player's prediction from his own rank is 2.13 whereas the absolute deviation of the second player's prediction from the 1st player's rank is 2.02 on average (Mann-Whitney test, $p = 0.710$). The first player's absolute deviation is 1.77 with respect to his revised prediction, and the difference between his revised prediction and the second player's prediction is insignificant (Mann-Whitney test, $p = 0.316$).

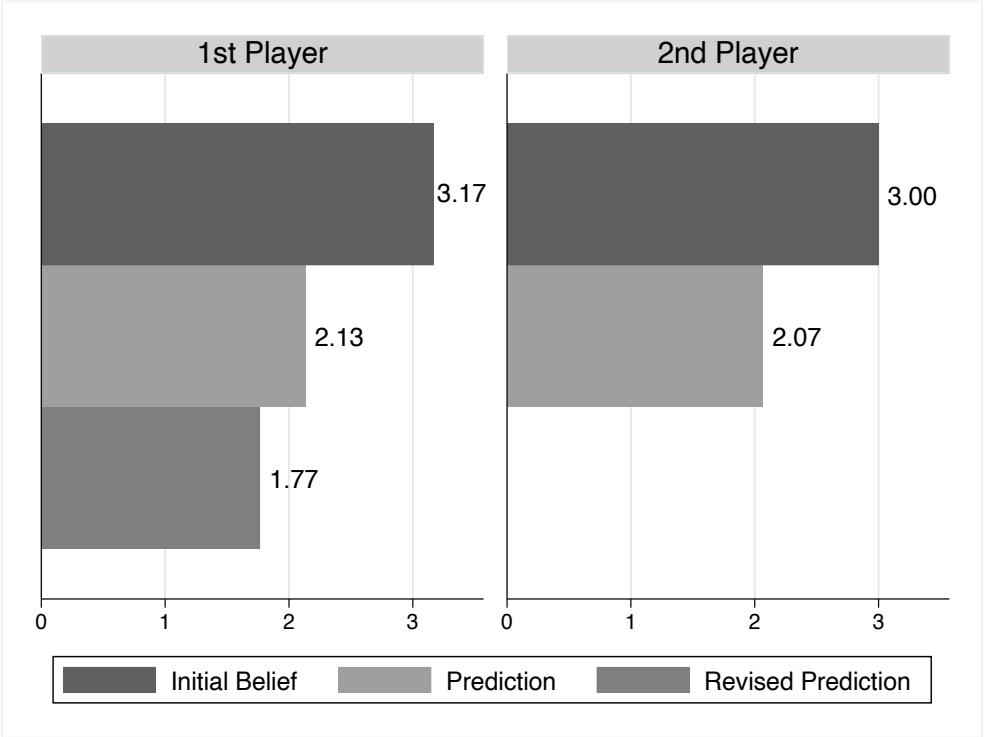
For the prediction of the second player's rank, we notice that the absolute deviation of the second player is 2.13 whereas the absolute deviation of the first player's prediction is 2.07 on average (Mann-Whitney test, $p = 0.720$). On the other hand,

the first player's revised prediction deviates from the 2nd player's rank by 2.05 in absolute values. The difference between the deviation from the 1st player's revised prediction and the 2nd player's prediction is not significant (Mann-Whitney test, $p = 0.819$).

C.4 Comparison of the Absolute Deviations from Own Rank

Figure C.8 shows the mean absolute deviation of the players' prediction about their rank from their actual rank for the 1st and 2nd player version, respectively.

Figure C.8: Mean Absolute Deviation from Player's Own Rank



We start with the prediction of the 1st player's rank. The absolute deviation of the first player's initial belief is 3.17 while the absolute deviation of the first player's prediction upon receiving a private signal is 2.13 on average (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). The first player's absolute deviation is 1.77 with respect to his revised prediction, and the difference between his revised prediction

and the previous prediction is also significant (Wilcoxon matched-pairs signed-rank test, $p = 0.022$). Overall, we can conclude that the 1st player makes more accurate predictions about his own rank as he receives a new piece of information.

When it comes to the prediction of the second player's rank, we notice that the absolute deviation of the second player's initial belief is 3.00 whereas the absolute deviation of the second player's prediction is 2.07 on average (Wilcoxon matched-pairs signed-rank test, $p = 0.000$). Similar to the first player, we can conclude that the 2nd player makes more accurate predictions about her own rank as she receives an information about her rank.

C.5 Instructions[1]

Welcome!

You are about to participate in an experiment in decision-making funded by universities and research agencies.

In this experiment, we first ask you to read instructions explaining the decision scenarios that you will face. Then, in the experiment itself, you will make decisions.

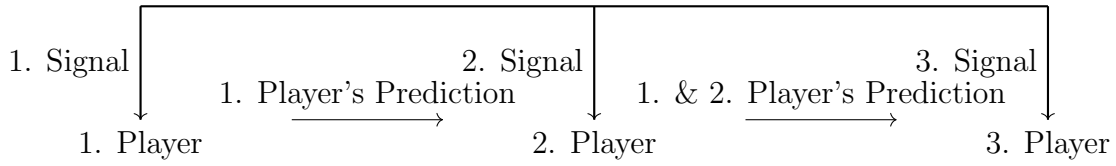
In the course of the experiment, you will earn money. Your monetary earnings will be determined by your decisions as well as by chance.

Simply for coming here and completing the experiment, you will receive a fixed participation fee of 10 Turkish Lira. Other earnings during the experiment will be in tokens currency. 10 tokens will be equal to 1 TL. All earnings that you make during the experiment will be added to the participation fee amount.

The experiment includes 5 sections and a brief survey. The experiment will last on average 1 hour.

Thank you.

C.5.1 Predict Number



Welcome to the first section of the experiment!

We will call this section PREDICT NUMBER for convenience. PREDICT NUMBER includes 2 parts. This section consists 6 rounds: the first 3 round is with partial information and last 3 round is with full information. In this section of the

[1]Original instructions were in Turkish and are available upon request. The instructions given here includes the treatment called the prediction on player 1's ranking with partial information under the section titled, *predict picture test ranking*. The other versions of the experiment includes the prediction on player 1's ranking with full information, player 2's ranking with partial information and player 2's ranking with full information treatments, respectively. These versions are left out for convenience and available upon request.

experiment your goal is to predict a target number. The target number is a randomly chosen number that is between 0 and 100.

At the start of this section of the experiment, the computer will randomly allocate all participants into groups of 3 player but your payoff is affected by only your choices. In each round, the groups will be reformed randomly.

Every player in each group will receive an independent private signal. The signals are in the form of comparison between the target number and a random integer between 0 and 100. For instance, the examples of signal can be the following if the target number is 60:

- Target number is greater than 30
- Target number is less than 90

Every player other than the 1stplayer in each group will receive an information regarding the decisions of the previous players in their groups. This information can be partial or full. If partial information is provided, the players will be informed that the previous players predicted that the target number is either larger than 50 or less than 50. For example, if the 1stplayer predicted 40 then the 2ndplayer will see the following in the partial information case:

- The 1stplayer predicted that the target number is less than 50

In the full information case, the players are given the precise prediction of the previous players. For instance, if 1stplayer predicted 40, 2ndplayer will see the following:

- The 1stplayer predicted that the target number is 40

Part I

Welcome! This is the first stage of PREDICT NUMBER.

In this stage, all participants will be randomly allocated into groups that includes 3 player. This section consists 3 round. Please note that groups will be reformed randomly at each round.

In this stage of the experiment your goal is to predict a target number. The target number is a randomly chosen number that is between 0 and 100. Please note that the computer will randomly choose a number that can take any value between the interval of 0 and 100. At each round, the target number will be randomly re-chosen by computer.

You can take the role of Player 1, Player 2 or Player 3 in each round. The respective stage will proceed as follows:

- 1-) Player 1 will get a private signal to make a prediction. Please note that this signal can be seen by only Player 1.

- 2-) Player 2 will receive full information about the Player 1's prediction in addition to an independent private signal to make a prediction. Please note that this new independent signal can be seen by only Player 2.

- 3-) Finally, Player 3 will receive full information about Player 1 and Player 2's prediction in addition to an independent private signal to make a prediction.Please note that this signal can be seen by only Player 3.

Your payoff from this stage is inversely related to the distance between your prediction and actual target number. the formula for the payoff in tokens from this section is the following:

$$100 - (prediction - targetnumber)^2/100$$

Your payment from this section will be determined by choosing one of the rounds randomly by rolling a 3-sided die. Therefore, each round is likely to be chosen and you should treat each round as the payment round.

Part II

Welcome! This is the second and last stage of PREDICT NUMBER.

In this stage, all participants will be randomly allocated into groups that includes 3 player. This section consists 3 round. Please note that groups will be reformed randomly at each round.

In this stage of the experiment your goal is to predict a target number. The target number is a randomly chosen number that is between 0 and 100. Please note that the computer will randomly choose a number that can take any value between the interval of 0 and 100. At each round, the target number will be randomly re-chosen by computer.

You can take the role of Player 1, Player 2 or Player 3 in each round. The respective stage will proceed as follows:

- 1-) Player 1 will get a private signal to make a prediction. Please note that this signal can be seen by only Player 1.

- 2-) Player 2 will receive partial information about the Player 1's prediction in addition to an independent private signal to make a prediction. Please note that this new independent signal can be seen by only Player 2.

- 3-) Finally, Player 3 will receive partial information about Player 1 and Player 2's prediction in addition to an independent private signal to make a prediction.Please note that this signal can be seen by only Player 3.

Your payoff from this stage is inversely related to the distance between your prediction and actual target number. the formula for the payoff in tokens from this section is the following:

$$100 - (prediction - targetnumber)^2/100$$

Your payment from this section will be determined by choosing one of the rounds randomly by rolling a 3-sided die. Therefore, each round is likely to be chosen and you should treat each round as the payment round.

C.5.2 Cyclical Predict Number

Welcome to the second section of the experiment!

We call this section CYCLICAL PREDICT NUMBER for convenience. In this section, all participants will be randomly allocated into groups that includes 2 player. This section consists 4 rounds: the first 2 round is with partial information and last 2 round is with full information. At the start of this section of the experiment, the computer will randomly allocate all participants into groups of 2 player but your payoff is affected by only your choices. In each round, the groups will be reformed randomly.

Part I

Welcome to the first stage of CYCLICAL PREDICT NUMBER!
In this stage, all participants will be randomly allocated into groups that includes 2 players. This section includes 2 round. Please note that groups will be reformed randomly at each next round.
In this stage of the experiment your goal is to predict a target number. The target number is a randomly chosen number that is between 0 and 100. Please note that the computer will randomly choose a number that can take any value between the interval of 0 and 100. At each round, the target number will be randomly re-chosen by computer.
You can take the role of Player 1 or Player 2. The respective stage will proceed as follows:

- 1-) Player 1 will get a private signal to make a prediction. Please note that this signal can be seen by only Player 1.

- 2-)Player 2 will receive partial information about the Player 1's prediction in addition to an independent private signal to make a prediction. Please note that this signal can be seen by only Player 2.

- 3-) Player 1 will receive partial information about Player 2's prediction. Then, Player 1 will make a revised prediction.

Your payoff from this stage is inversely related to the distance between your prediction and actual target number. the formula for the payoff in tokens from this section is the following:

$$100 - (prediction - targetnumber)^2/100$$

Your payment from this section will be determined by choosing one of the rounds randomly by a coin flip. If you are Player 1 for the chosen round, one of your two predictions will be chosen by another coin flip. Therefore, each round is likely to be chosen and you should treat each round as the payment round.

Part II

Welcome! This is the second and last stage CYCLICAL PREDICT NUMBER.
In this stage, all participants will be randomly allocated into groups that includes 2 players. This section consists 2 round. Please note that groups will be reformed randomly at each next round.
In this stage of the experiment your goal is to predict a target number. The target number is a randomly chosen number that is between 0 and 100. Please note that the computer will randomly choose a number that can take any value between the interval of 0 and 100. At each round, the target number will be randomly re-chosen by computer.
You can take the role of Player 1 or Player 2. The respective stage will proceed as follows:

- 1-) Player 1 will get a private signal to make a prediction. Please note that this signal can be seen by only Player 1.

- 2-)Player 2 will receive full information about the Player 1's prediction in addition to an independent private signal to make a prediction. Please note that this signal can be seen by only Player 2.

- 3-) Player 1 will receive full information about Player 2's prediction. Then, Player 1 will make a revised prediction.

Your payoff from this stage is inversely related to the distance between your prediction and actual target number. the formula for the payoff in tokens from this section is the following:

$$100 - (prediction - targetnumber)^2/100$$

Your payment from this section will be determined by choosing one of the rounds randomly by a coin flip. If you are Player 1 for the chosen round, one of your two predictions will be chosen by another coin flip. Therefore, each round is likely to be chosen and you should treat each round as the payment round.

C.5.3 Raven's Progressive Matrices

Welcome to the third section of the experiment!
The following section will include a series of questions. We call this section PICTURE TEST for convenience. In each question, there is a picture with a piece missing . Your goal is to find the missing piece for each picture. Please choose the answer that shows the piece that completes the picture. For each correct answer, you will be rewarded 10 tokens. You need to solve 23 questions in 10 minutes. You can find an example below:

C.5.4 Predict Picture Test Ranking

Welcome to the fourth and PREDICT PICTURE TEST RANKING part of the experiment!
In this section, all participants will be randomly allocated into groups that includes 2 player. This section consists 2 rounds. Groups will be randomly reformed at each round.

If you are Player 1 in the 1stround, then you will be Player 2 in the next round; vice versa. Your goal is to predict Player 1's ranking (target ranking) in the previous

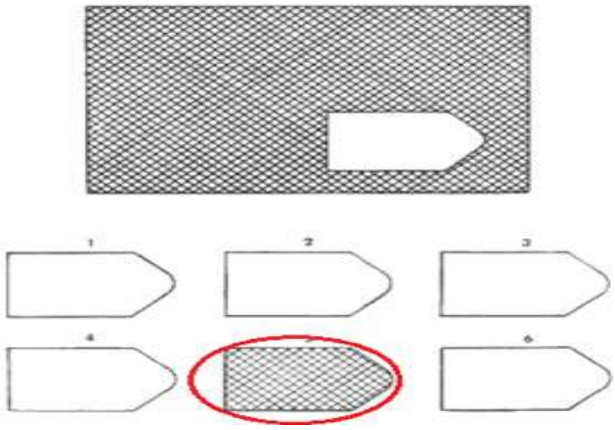


Figure C.9: Picture Test Example

test where 1 is the best possible rank and 12 is the worst possible participant rank. Ranks are calculated according to the number of correct answers on the picture test. Please note that if you are Player 1 in a round, then you should predict your own ranking whereas if you are Player 2 in a round, you will be predicting the other group member's ranking in the test.

Each player will get a private independent signal to make a prediction. The signal will be in the form of comparison between a random participant's ranking and target ranking.

For instance; if the target ranking (Player 1's ranking) is 3, then signal can be:

- Player 1's ranking is better than 7th rank.
- Player 1's ranking is worse than 1strank.

The steps for this section is as follows:

- Player 1 will state its belief about its test ranking.
- Player 1 will get a private signal to make a prediction. Please note that this signal can be seen by only Player 1.

- Player 2 will receive partial information about the Player 1's prediction. Partial information is such that the prediction of other player in your group is either in the top half (between 1 and 6) or in the bottom half (between 7 and 12) of the ranking pool. Additionally, Player 2 will get an independent private signal to make a prediction. Please note that this signal can be seen by only Player 2.

- Player 1 will receive partial information about Player 2's prediction to make a revised prediction.

Your payoff from this stage is inversely related to the distance between your prediction and actual target number. the formula for the payoff in tokens from this section is the following:

$$150 - (prediction - targetranking)^2$$

Your payment from this section will be determined by choosing one of the rounds randomly by coin flip. If you are Player 1 in the payment round, one of your predictions will be chosen by another coin flip. Therefore, each round is likely to be chosen and you should treat each round as the payment round.

C.5.5 Risk Preference Elicitation

Welcome to the fifth and last section of the experiment!
In this section, you will be provided with two options that involve rewards depending on given probabilities.

For each line below, please choose whether you prefer Option A or Option B. Your payment from this section will be determined by your choice (Option A or Option B) from a randomly selected line. We will roll a 10-sided die to determine the line for the payment.

For example; if you choose option A for the first line, you will be awarded 100 tokens with probability of 1/10 and 80 tokens with probability of 9/10; whereas if you choose option B for the first line, you will be rewarded with 190 tokens with probability of 1/10 and 5 tokens with probability of 9/10. Each choice could be the one that counts, so you should treat each and every line as if that choice will determine your payment.

Table C.1: Risk Preference Elicitation

Option A				Option B				Option A	Option B
Prob	Payoff	Prob	Payoff	Prob	Payoff	Prob	Payoff		
1/10	100	9/10	80	1/10	190	9/10	5	☐	☐
2/10	100	8/10	80	2/10	190	8/10	5	☐	☐
3/10	100	7/10	80	3/10	190	7/10	5	☐	☐
4/10	100	6/10	80	4/10	190	6/10	5	☐	☐
5/10	100	5/10	80	5/10	190	5/10	5	☐	☐
6/10	100	4/10	80	6/10	190	4/10	5	☐	☐
7/10	100	3/10	80	7/10	190	3/10	5	☐	☐
8/10	100	2/10	80	8/10	190	6/10	5	☐	☐
9/10	100	1/10	80	9/10	190	1/10	5	☐	☐
10/10	100	0/10	80	10/10	190	0/10	5	☐	☐

C.5.6 Post-experiment Survey Questions

Please provide the information requested below, but do not write your name. (You can be assured that all information will be stored in a 100% anonymous way to ensure your privacy.)

- How old are you?:
- What is your gender?:

- What is your major?:

- Graduate/undergraduate year of study?: