UNIVERSITÄT HAMBURG

DISSERTATION

Institutional Design for Cooperation: Essays on Economic Experiments and Applied Microeconomic Theory

---

Zur Erlangung der Würde einer Doktorin der Wirtschafts- und Sozialwissenschaften Dr. rer. pol.
(gemäß der Promotionsordnung vom 18. Januar 2017)

vorgelegt von
Fanny Elise Schories
aus Karlsruhe

Oktober 2021

Erstgutachter: Prof. Dr. Gerd Mühlheußer
Zweitgutachter: Prof. Dr. Stefan Voigt

Datum der Disputation: 30. März 2022

Für Renate

Acknowledgements

Writing this dissertation would not have been possible without the help and support of a number of people. First and foremost, I would like to thank my supervisor Gerd Mühlheußer, who was always ready to provide me with encouragement and patience and from whose advice I have benefited greatly.

I am grateful to my second supervisor, Lydia Mechtenberg, for many helpful comments and ideas.

A very special thanks is due to Stefan Voigt and the entire Institute of Law and Economics, especially Nora El-Bialy, who enabled me to explore many new places – academically and geographically – in the DFG-Graduate School "The Economics of the Internationalization of the Law" and in the DAAD-Project "Dialogue with the Muslim World".

I had the great fortune to work with two amazing co-authors, Christos Litsios and Stephan Michel, whose enthusiasm and intelligence are unrivaled. In addition, all my colleagues at the Graduate College were an integral part of making the PhD journey insightful and fun.

Sometimes a change in scenery works wonders, too. I am very thankful to James Andreoni for inviting me to the University of California in San Diego and to Rania Miniesy and Dina Rabie for hosting me at the British University in Egypt.

Financial support of the German Research Foundation (GRK 1597/3), the German Academic Exchange Service, and the WiSo Graduate School enabled the pursuit of my projects and are gratefully acknowledged.

I will forever be grateful for the unconditional and cheerful support of my family, especially Renate Schwarzbauer, Hartmut Schories, and Bernd Osterhus. Finally, I thank Anna Both and Felix Hadwiger, my two favorite people.

Contents

Introduction

No man is an island,
Entire of itself,
Every man is a piece of the continent,
A part of the main.
John Donne, 1624

Achieving Cooperation: A Long-Lasting Struggle Nature is brimful of examples where cooperative behavior increases overall well-being: bats feeding other colony members, bees giving their own lives to defend the hive, insects and plants entering into irreversible symbioses. Likewise, much of humankind's history is shaped by the struggle to achieve sustainable cooperative agreements as "many of the benefits sought by living things are disproportionally available to cooperating groups" (Axelrod and Hamilton, 1981).

Situations of a structure where cooperation between actors creates a mutual benefit, but defection reaps individual gains at others' expense are subsumed under the term social dilemma. Social dilemmas receive high interest from social science researchers across disciplines. Unsurprisingly so, as the insights can be applied to countless well-known situations in everyday life, from geopolitics to competition law and dormitory cleaning rosters. Pareto-efficient outcomes in a social dilemma can be reached if agents are motivated to pursue a common interest instead of individual gains. But, alas, the incentive to defect from cooperation is a constantly luring temptation. How can it be tamed?

Consider the institutions that shape the interactions between humans: the legal and social "rules of the game" (North, 1990). Institutions provide the framework in which actors make their individual decisions. In a game-theoretic sense, they define – and can therefore also re-define – the payoff structure that agents face. Institutions may alter the incentives of the game at play in a way that makes cooperative behavior more attainable. Beyond providing formal rules and procedures, a recurring theme in this dissertation is the argument that institutions affect cooperative behavior in more subtle ways than can be captured in monetary terms (Ostrom, 1998).

The papers in this dissertation consider various institutional arrangements that all share the common goal of investigating the factors influencing cooperation between players. The studies represent a continuum regarding the strictness with which the institutional design is able to enforce cooperation. Therefore, varying

aspects of formal and informal institutions are considered. The first three chapters of the dissertation share a common method: economic laboratory experiments. Economic experiments are a powerful tool to observe the effects of institutions on humans, whose ways are often more intricate than the *homo oeconomicus* model suggests. Experiments offer a controlled environment to study how human subjects respond to institutional settings. The controlled environment allows us to test hypotheses by varying the treatment variable of interest while keeping crucial confounding factors – information, personal characteristics, and preferences – fixed. The fourth chapter develops a theoretical model to answer a question that is to some extent normative: how *should* institutions be designed to best meet the goals of the designers?

In the following, I elaborate on the specific aspects of institutional design and cooperation covered by the papers that form this dissertation. I present each chapter individually to briefly discuss the specific research questions, the limitations and added value of the studies, and then draw some general conclusions.

Chapter 1: Image Concerns Discipline Selfish Bargainers The first chapter investigates how the institutional environment can push behavior towards fairness in distributional decisions. We show that transparency is a powerful tool to activate generosity and equity considerations even in selfish individuals by stimulating image concerns. Subjects are more willing to share with their counterparts if the decision is observed by others, and are moreover less willing to invest into acquiring bargaining power. Different kinds of institutions interact in this setting: by ensuring transparency, the formal institutional structure shapes the exchange here, but it is the informal institution of a fairness norm that dictates what behavior is deemed desirable.

At the heart of the experiment is a distributional decision modeled by the dictator game (Kahneman, Knetsch, and Thaler, 1986). In this game, the dictator freely decides how to divide a sum of money between herself and a passive recipient. In order to assess how transparency affects the giving decision and the preference to be in the role of the dictator, we systematically vary whether information on the giving decision is revealed to other subjects. The preference intensity over the roles is elicited via a second-price auction. Our experimental design thus goes beyond the pure sharing decision: not only are we the first to explicitly measure role-preferences in this game, but we are also able to asses these in interaction with image concerns.

Adding the auction in interaction with the publicity treatment creates two important insights: First, subjects are willing to spend considerable amounts of their endowment in order to become the dictator – (correctly) anticipating that this role is the one resulting in the higher payoff. But since the money spent in the auction for the dictator role is withdrawn from the game, bidding high decreases the joint welfare of the pair. Second, stimulating image concerns increases the amount shared and lowers the willingness to pay for the role and the bargaining power of a dictator.

Hence, transparency increases efficiency and equity in situations where social norms prescribe pro-social behavior.

The relevance outside the lab spans further than interactions between individuals only. For example, if politicians and government authorities are required to report to their constituents about the effectiveness of regulation in specific areas, they can be expected to exert more effort on these projects because their public image is at stake. In the area of corporate social responsibility, companies will be more likely to implement equal pay and fair conditions along their supply chains if this information is made publicly available.

The literature on nudging is built around the fact that small changes to a choice architecture lead individuals to change their behavior even when a selfish *homo oeconimucs* would not: in the absence of direct financial repercussions, a rational actor considering only their own monetary payoff would not change behavior due to increased transparency. We demonstrate that this is not the case: image concerns are a powerful motivator. The first chapter thus shows how pro-social behavior can be attained even with minimal formal institutions when actors are motivated by non-monetary objectives.

Representative Democracy and Cooperation Chapters 2 and 3 delve deeper into the influence that democratic institutions and decision-making procedures exert onto cooperative behavior. The study of how to solve cooperation problems is particularly relevant in the political realm: the existence of social dilemmas is the central argument for ceding authority to the state in order for it to enforce welfare-improving outcomes (Ostrom, 1998). A topical application of the insight that formal institutions are a necessary but not sufficient condition for a functioning democratic state is discussed in "How Democracies Die" by Levitsky and Ziblatt (2018). The authors assert that a democracy cannot survive in the long term without a shared sense of political legitimacy and voluntary cooperation, since constitutions are essentially incomplete contracts. No set of written rules can possibly identify and address all aspects of a political system, thereby leaving legal loopholes for egocentric actors to exploit. To create stable democratic governance under these circumstances requires sustaining a shared understanding of legitimacy and trustful cooperation. This idea is echoed in the experiments with a focus on representative democracy that are presented in the second and third chapter. They present evidence how the way in which institutions are chosen influences cooperative behavior, even when the formal incentives are held constant. The increase in subjects' cooperation in response to participatory procedures is called *democracy premium* (Dal Bó, Foster, and Putterman, 2010). The following paragraphs present Chapters 2 and 3 individually.

Chapter 2: Elected Leaders Induce More Cooperation Than Unelected Ones The second chapter examines whether an institution has a differing impact on cooperation if it is introduced by a representative of the affected parties rather than exogenously imposed. It is shown that the way a leader is chosen matters for the willingness of subjects to behave cooperatively under the institutional setting imposed by the leader: elected representatives are much more effective in inducing cooperation than randomly appointed leaders. The study thus contributes to the literature on the effects of endogenous institutions by focusing on the aspect of representation.

The experimental treatment varies whether a decision-maker is democratically elected or randomly appointed. There is evidence of a large democracy premium – more cooperative behavior – only if the group leader is democratically chosen. For randomly appointed leaders, no democracy premium can be found. Especially the subjects who initially did not prefer the policy are more likely to cooperate if it was brought about by an elected representative, pointing out the importance of decision-making procedures carrying democratic legitimacy.

The insights speak to any body with the authority to make decisions on behalf of others: elected leaders, ceteris paribus, embody more legitimacy than unelected leaders, which in turn affects compliance with their decisions. This is directly applicable to the division of labor between the legislative and executive government branches. Consider, for instance, the implementation of the regulation for the containment of the Coronavirus pandemic in Germany: A large discussion arose around the question whether acts should be enacted by the (elected) parliament or the (appointed) cabinet.[1] The study in this chapter draws on rigorous economic experimental methodology to offer a practical argument for the involvement of the parliament: citizens can be expected to comply with rules more if elected representatives were the ones enacting them.

Chapter 3: The Influence of Democracy on Cooperation Is Not Universal The third chapter approaches the behavioral implications of institutions from a comparative perspective and shows that they depend on culture and context. The experiment on the effects of representative democracy is conducted at universities in Cairo and in Hamburg. The results show that reactions to democratic procedures are influenced by political culture. In Germany, the democracy premium appears large and significant. In Egypt, no democracy premium can be found.

Drawing on the setup developed in the previous chapter, the behavioral response to representative democracy is tested and complemented with questions from the World Value Survey by Inglehart et al. (2014) to asses the transmission channel between societal values and responses to democracy. By comparing the experimental results from Egypt and Germany, I find that the aforementioned impact of democratic decision-making procedures is not universal. In Egypt, on the

[1] See for example statements by the Federal Constitutional Court's president, Stephan Harbarth that – contrary to the actual modus operandi of the government – all crucial decisions should be made in the parliament (mimeo, 2021).

one hand, strong preferences for democratic participation were made vocal and visible during the Arab spring. On the other hand, numerous authors doubt whether the institutional framework provided by Islam is compatible with democracy at all (Fukuyama, 1992; Huntington, 1996; Kumar, 2010). My study does not establish a positive connection between representative democracy and cooperation on the micro-level for Egyptian students. But the data provides no evidence of religiosity as the determining factor. Contrary to popular hypotheses, religiosity and adherence to Islam is not found to be negatively related to preferences for democracy. Instead, trust levels and obedience towards authority seem to be driving the results. I show that trust in a society is an important complement to the formal institutions provided by a representative democracy.

Chapter 4: Institutional Interactions Matter for International Cooperation The fourth chapter zooms out further to study how states can form cooperative agreements in the form of an international organization (IO). We theoretically investigate the institutional design choices that countries face when trying to provide a club good together. We are the first to explicitly model the interaction between the two main design choices an IO faces: the voting rule and the accession costs charged from new member states. We use a screening model to investigate how an IO might balance the benefits of enlargement with the risk of becoming less effective in club good provision whenever there is uncertainty about the productivity of candidate states. The results show that the voting rule is an efficient screening device, but incumbent states may prefer to charge accession costs in order to extract rents.

The example discussed in the chapter is the formation of a common currency area like the Eurozone, but the results speak to a broader set of problems in which cooperation between countries creates mutual benefits: especially the current crises of the Coronavirus pandemic and global warming are issues of such a kind where incentives to free-ride on others' efforts loom large: without effective institutional governance, individual countries are forever tempted to ban the export of vaccines in times of scarcity and to reduce their greenhouse gas emissions less than other countries. A solution is to change the institutions and thus incentives of the interaction, for example to turn the production of a public good into a club good. A club has the advantage that any non-contributor can be excluded from the benefits and free-riding is not a profitable option anymore. This is where the fourth chapter moves further along the spectrum of institutional design to address cooperation problems: whereas the institutions in the previous chapters relied on voluntary compliance, cooperation becomes enforceable in a club. Nordhaus (2015) discusses the incentive transformation towards a club for the case of climate change: by linking trade policies to efforts to reduce carbon emissions, a carbon tariff makes it incentive-compatible for countries to increase their efforts in tackling climate change. Our results are able to inform the specific choices that need to be addressed in the formation of such a club and make the agreement more successful.

What Can Be Learned From The Results? To summarize, the central topic of this dissertation is the notion of cooperative behavior in non-cooperative games and how it is shaped by institutions. Cooperation is a central concern in today's political debates: states are called to cooperate to cope with the challenges of climate change; citizens are called to cooperate with public health regulations and the governance of common goods. Usually, in such cases from everyday life formal institutions are present but cannot regulate every minute detail of behavior: e.g., in the presence of private information or due to a lack of enforcement mechanisms. What the four chapters demonstrate is that great attention has to be paid to the details of institutional design if the goal is to attain cooperation between (political) actors. In a nutshell, first, image concerns influence preferences over the role assignment in a distributional decision and increased transparency can lead to more equitable allocations. Second, small changes which preserve the formal incentive structure of the game, nevertheless influence cooperative behavior and overall welfare. Attention has to be paid to the perceived legitimacy of a process. Third, (political) culture and trust in a society matter for the ability of decision-making procedures to foster cooperation. Fourth, institutional design features have to be considered in their interaction.

Bibliography

Axelrod, Robert and William D Hamilton (1981). "The Evolution of Cooperation". In: *Science* 211.4489, pp. 1390–1396.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy". In: *American Economic Review* 100.5, pp. 2205–29.

Fukuyama, Francis (1992). *The End of History and the Last Man*. New York: Avon.

Huntington, Samuel (1996). *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster.

Inglehart, R, C Haerpfer, A Moreno, C Welzel, K Kizilova, J Diez-Medrano, M Lagos, P Norris, E Ponarin, and B Puranen (2014). *World Values Survey: Round Six - Country-Pooled Datafile Version*.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1986). "Fairness and the assumptions of economics". In: *Journal of Business*, S285–S300.

Kumar, Deepa (2010). "Framing Islam: The resurgence of Orientalism during the Bush II era". In: *Journal of Communication Inquiry* 34.3, pp. 254–277.

Levitsky, Steven and Daniel Ziblatt (2018). *How Democracies Die*. New York: Crown.

mimeo (2021). "Grundrechte sind keine Privilegien". In: *Rheinische Post*. URL: `https://rp-online.de/politik/deutschland/praesident-des-verfassungsgerichtes-grundrechte-sind-keine-privilegien_aid-56152945` (visited on 06/20/2021).

Nordhaus, William (2015). "Climate clubs: Overcoming free-riding in international climate policy". In: *American Economic Review* 105.4, pp. 1339–70.

North, Douglass C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Ostrom, Elinor (1998). "A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997". In: *American Political Science Review*, pp. 1–22.

Chapter 1

Role Preferences and Social Image Concerns in the Dictator Game

Authors Christos Litsios, Fanny Schories

Abstract We examine role preferences in interaction with social image concerns in a dictator game experiment. The role assignment within subject pairs in the experiment is endogenous: subjects bid for the dictator role in a second-price auction. The dictator can then freely implement a division of a given amount of surplus between herself and the recipient. We measure the willingness to pay for the dictator role in two settings: when the payout for the game is made in private and when it is made in public. Our hypotheses follow the arguments of Bénabou and Tirole (2006) and Andreoni and Bernheim (2009) that the offer made by the dictator is a signal that reveals information about her fairness type, a motive which is amplified in the public treatment. We find that role preferences and image concerns are indeed connected: the willingness to pay for the dictator role decreases and the dictator offer increases when image concerns are activated.

Keywords Laboratory Experiment, Dictator Game, Social Image, Auction

JEL Classification A13, C91, D44, D63, D91

1.1 Introduction

Are people willing to forgo a strategically advantageous bargaining position in order not to be perceived as greedy? We investigate how the trade-off between monetary interests and social image concerns affects first-mover preferences in a dictator game experiment. In the dictator game – one of the workhorses of experimental research in economics on social preferences (Kahneman, Knetsch, and Thaler, 1986; Forsythe et al., 1994) – one player (the dictator) divides a monetary surplus between herself and another player (the recipient). In most experimental settings, subjects are exogenously matched in pairs, and each player is assigned one of the two roles at random. From experimental evidence and game-theoretical considerations, one easily deduces the dictator role to be the more advantaged role from a monetary perspective.[1] An experimental subject in the position to choose between the dictator and recipient role is therefore expected to strongly prefer the dictator role. In this paper, we show via an auction that changing the observability of the dictators' identity significantly affects preference intensities over the two roles.

Social interaction is shaped by norms and expectations of appropriate behavior that counterbalance the human base motives of greed and envy. For example, the principle of *pay what you want* (PWYW) has spawned a pricing strategy around the notion that social context sufficiently motivates customers to voluntarily pay for a product that they could technically consume for free. Distributional decisions such as paying for a beer at a PWYW bar, taking the last piece of cake from the office kitchen, and donating to a fundraiser collection are arguably all influenced by the social context of the situation. Especially the question "Is someone watching?" can powerfully stimulate image concerns. Giving a generous tip may cause some pain to a frugal individual, but not as much as being perceived as greedy by a potential partner in business or romance. The experiment presented in this paper systematically varies the observability of a sharing decision to quantify the influence of social image concerns on sharing behavior. Furthermore, the experimental set-up enables us to gauge the readiness by which people avoid entering a situation that invokes the trade-off between image and monetary concerns in the first place.

So far, the role preference intensity in dictator games has not been explicitly investigated, even less so in combination with image concerns. Previous research found ambiguous results regarding the influence of observability on sharing decisions. The most closely related studies are Dana, Cain, and Dawes (2006), Dufwenberg and Muren (2006), and Andreoni and Bernheim (2009). Dana, Cain, and Dawes (2006) demonstrate that subjects are willing to leave money on the table in order to escape the dictator role. But Dufwenberg and Muren (2006) find that public payouts make dictators *less* generous. In contrast, Andreoni and Bernheim (2009) show how image concerns make dictators more likely to split the surplus equally in a signaling

[1] See Camerer (2003) and Engel (2011) for meta-studies of the dictator game, where it is found that dictators share on average 20 to 30% of the pie with the recipient.

model of the dictator game and support the theoretical argument by experimental evidence. The contribution of our paper is to shed further light on the ambiguous effect of image concerns in combination with a quantitative measure of the role preference. Our hypothesis follows Andreoni and Bernheim (2009): taking on the role of the dictator inflicts additional (non-monetary) costs which are increasing in the visibility of a player's decision. By proposing an allocation, a player sends a signal about herself to the audience. She incurs a trade-off between securing a high monetary payoff and wanting to be perceived as fair. An equitable dictator offer signals a high intrinsic level of virtue and corresponds with a favorable image. Low offers signal selfishness and correspond with a poor image. Observability of the dictator's action thus reinforces image concerns and thereby decreases the expected dictator payoff, which translates into a weaker preference for this role.

The experimental design is as follows. First, we elicit all subjects' fairness preferences using a dictator game played via the strategy method (DG 1). Next, we match subjects in pairs and partially endogenize the role assignment for the main dictator game (DG 2). In a second-price auction, both players bid to increase the chances of becoming the dictator in DG 2. Winning the auction increases one's chances to become the dictator to 90 percent. The actual role assignment is done via a lottery to prevent players from deducing the result of the auction from their role with certainty. This way there remains some uncertainty with respect to the results of the auction and potential effects of entitlement, over which mixed evidence has been found (Hoffman et al., 1994; Demiral and Mollerstrom, 2018), are mitigated. After the lottery, each subject pair has one dictator and one recipient. Subjects then play DG 2 as a regular one-shot dictator game in the assigned roles. At the end of the session, the *public* treatment variation requires every dictator to stand in front of all participants of that session while their sharing decision is publicly announced and the recipient can identify her. In the *private* treatment, the dictators' decision is kept private, which is equivalent to the usual protocol of dictator experiments.

As hypothesized, we find that dictator offers are higher in the public treatment. The main driving factors are own fairness concerns as well as awareness of one's own image concern. Correspondingly, the willingness to pay to become the dictator is significantly lower in the public treatment compared to the private treatment. Here again, own fairness and awareness of one's own image concern explain the results. In addition, expectations about others' fairness increase the value of the recipient role and thereby decrease bids in the private treatment. With the public announcement, this effect vanishes as the general expected generosity shifts upwards.

The paper proceeds as follows. Section 2 discusses the most closely related literature. Section 3 presents the experimental design, an adaptation of the model by Andreoni and Bernheim (2009), and our hypotheses. The data analysis and results are shown in Section 4. Section 5 discusses the results and concludes.

1.2 Related Literature

Several papers have gradually reduced the dictators' anonymity to gauge the influence of image concerns on giving decisions. Hoffman et al. (1994) and Hoffman, McCabe, and Smith (1996) vary the degree of anonymity between the dictator and the experimenter. They find that offers increase as dictators' anonymity towards the experimenter decreases, showing that it is not so much a taste for fairness itself that makes people share in bargaining games, but rather a social concern for what others may think of them. Frey and Bohnet (1995) and Bohnet and Frey (1999) systematically vary the social distance between dictator and recipient. In their experiments, subjects stand up and look at other participants *before* playing dictator games. While the approach introduces image concerns as well, it crucially differs from our procedure with regard to the timing of information provision. In their set-up, dictators gather information about their opponent before making a choice. Image concerns and fairness preferences are likely to vary with respect to personal characteristics such as gender, ethnicity, age, et cetera. We believe our ex-post revelation procedure to be more conservative when it comes to separating image concerns from other confounding factors. With the ex post publication of the decision we are able to retain recipients' anonymity while maximizing the dictators' visibility. A more closely related experimental design is presented by Dufwenberg and Muren (2006), who make the payout of the dictator game either in front of an entire classroom or privately in an office, a procedure with high similarities to the one presented in this paper. The authors find that significantly less is shared when the payment is made in public. However, Dufwenberg and Muren (2006) cannot exclude that the effect is driven by the specific subject pool consisting of undergraduate economics students, who arguably follow different norms and expectations of how a dictator ought to behave, especially in the presence of the instructors teaching them game theory. As our sample consists of less than one-third economics and business students and the experiment is played outside the context of a classroom and lecture, we do not expect these experimenter-demand factors to have a significant influence in our case.

Gächter and Fehr (1999) and Rege and Telle (2004) use experimental designs similar to ours in order to increase visibility and hence social control in a distributional decision: subjects' contributions to a public good are made public to all participants of a session. The former find no effect on contributions, the latter a positive effect. The occurrence of mixed or even ambiguous results is a recurring phenomenon in the literature on social cues and generosity. Even though the overall insight from (Engel, 2011)'s meta-study is that there is a positive relationship between the two, some studies find null (Laury, Walker, and Williams, 1995; Bolton, Katok, and Zwick, 1998; Gächter and Fehr, 1999; Johannesson and Persson, 2000; Frohlich, Oppenheimer, and Moore, 2001; Barmettler, Fehr, and Zehnder, 2012; Dreber et al., 2013) or even contradictory results (Dufwenberg and Muren, 2006; Rankin, 2006).

For our theoretical analysis we rely on work by Andreoni and Bernheim (2009),

who propose a simple inequity aversion model and extend it by social image concerns. Thereby, a player's behavior is not only driven by her own preference to act socially but also by the impression her action leaves on others. In such a signaling game, players increase their dictator offers. Andreoni and Bernheim (2009) explain pooling at the equal split of the surplus and show in an experiment how such pooling behavior is driven by image concerns. Our argument concerning role preferences in the dictator game builds on the signaling rationale introduced by Andreoni and Bernheim (2009). In contrast to their model we introduce image concerns as additional (non-monetary) costs instead of benefits. This assumption is in line with findings by Dana, Cain, and Dawes (2006), who show that subjects have a positive willingness to pay to avoid being in the situation of a dictator. Broberg, Ellingsen, and Johannesson (2007) estimate exit reservation prices in a modified replication of Dana, Cain, and Dawes (2006) and find that approximately two-thirds of the subjects are happy to quietly leave a dictator game with a smaller amount than the surplus to be divided. Andreoni, Rao, and Trachtman (2017) present a field experiment with similar findings. Our contribution in this line of research is to quantify the influence of social control on the willingness to pay for the dictator role.

To the best of our knowledge no experiment has used an auction to allocate the roles in a dictator game before. There are two experiments using auctions to sell participation with pre-defined roles in the ultimatum game: Güth and Tietz (1986) via a second-price auction and Shachat and Swarthout (2013) via an English clock auction. Subjects are randomly allocated to being either in the potential proposer or the potential responder pool and can then bid on entering the game. Güth and Tietz (1986) find that proposers bid on average almost twice as much as responders. Shachat and Swarthout (2013) find that auction prices often reflect beliefs inconsistent with Nash equilibria (nota bene of a game including only monetary payoffs). Our experiment differs insofar, as both aforementioned papers use an ultimatum game and are not able to measure the intensity of the preference for being the first-mover directly, since roles are predetermined.[2] Furthermore, the auction creates an entitlement effect because not every subject could participate in the following ultimatum game. Entitlement has been found to significantly influence behavior (Hoffman et al., 1994; Frohlich, Oppenheimer, and Kurki, 2004; García-Gallego, Georgantzís, and Jaramillo-Gutiérrez, 2008). We therefore interpose a randomization between the auction and the bargaining situation to minimize such entitlement effects: a proposer cannot know for sure how she got her position and thus should not feel a stronger entitlement of the surplus than the responder.

To conclude, the monetary attractiveness of the dictator role seems indisputable from a game-theoretic perspective, especially for selfish players. However, behavioral insights from other studies around social image and inequity aversion contest

[2]The willingness to pay for one role over the other can only be implicitly estimated as the difference in the willingness to pay between the potential proposers and responders and by assuming that the value of the game itself is constant across the two roles. The difference measure is not accurate once there are interaction effects between the game value and the role preference.

this presumption, which has not been rigorously tested so far. Our main treatment variable social image has been found to produce ambiguous results in previous studies, suggesting that additional systematic evidence is needed to shed light on the exact transmission mechanism between social context and bargaining behavior.

1.3 Experimental Design

1.3.1 The Games Played

The experiment consists of three parts as depicted in Figure 1.1. In Part 1, subjects play a dictator game (DG 1) via the strategy method not knowing if they will be paid as dictator or recipient. Every player makes the hypothetical choice how to distribute 40 points between herself and another unknown player. Which players are matched in pairs and whose allocation is implemented is randomly chosen. Subjects learn about the realized payoff from this part at the very end of the session and the payoffs from DG 1 are never made public. The strategy method allows us to use the number of points shared in the first dictator game as a measure of every subject's fairness preferences in the absence of image concerns arising from the public payment. During Part 1, subjects are aware that other parts will follow, but remain ignorant of the exact games to be played later on. All actions are taken in a one-shot manner. We elicit first-order beliefs of subjects in DG 1 in an incentivized way similar to Krupka and Weber (2013).

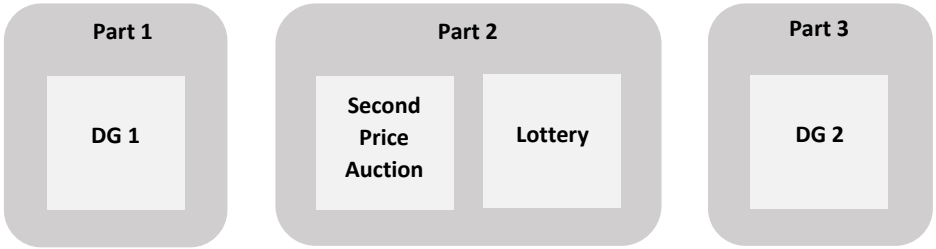


FIGURE 1.1: Experimental Set-Up

In Part 2, subject pairs are matched for the rest of the experiment. Subjects learn that another dictator game (DG 2) will follow. Both players $i = 1, 2$ of a pair privately place a bid in a sealed-bid second-price auction to increase the probability of becoming the dictator in the following way: They each receive an endowment of 100 points and make a bid $b_i \in [0, 100]$. The higher bid in a pair buys a lottery ticket, which wins this player the dictator role in DG 2 with 90% probability and the recipient role with 10%. Vice versa, the player with the lower bid wins the dictator role with a probability of 10% and the recipient role with 90%. The auction winner pays the other player's bid to the experimenter, the loser keeps her entire endowment from Part 2. We use this strategy-proof mechanism to gauge subjects' willingness to pay in an incentive-compatible way, as bidding one's true valuation is a weakly

dominant strategy. In case of a draw the auction winner is randomly determined by the computer with equal chances and pays her bid.

The lottery draw is carried out secretly by the computer and independently for each pair. At the end of Part 2, subjects are informed about the role assignment after the lottery, but neither learn who won the auction, nor the price the winner pays for the ticket. The lottery is meant to limit spillovers of confounding factors between the auction and DG 2. First, learning about the opponent's bid, and thereby her willingness to pay for the dictator role, could be informative of her player type. Second, winning the auction could trigger the aforementioned entitlement effects. Via the lottery we keep the players unaware of whether it was their bid that got them the role they are in, or sheer luck. However, a winning probability of 90% keeps the noise minimal and is still high enough to make bidding attractive.[3]

After the role assignment, the dictator game DG 2 is played in Part 3, this time with direct response instead of the strategy method.[4] The dictator splits an additional surplus of 100 points between herself and the recipient. The procedures of the experiment as described up to this point are the same in both treatments. Our treatment variation comes into play for the implementation of the dictator's allocation, as described in the next section.

1.3.2 The Public Treatment

We have two treatments conditions, a *public* and a *private* one. In the public treatment, at the very end of the session, dictators are required to step out of their cubicle such that all other participants can see them. The experimenter reads aloud the list of cubicle numbers along with the players' sharing decision made in DG 2. The recipients see their corresponding dictator's cubicle number and sharing decision on their screen and can thus identify the person they were paired with, as well as learn about all other dictators' decisions. The recipients do not stand up and are not identified. The instructions of the public treatment describe this procedure in detail to make the treatment salient to subjects at the time when they make their decisions in Parts 2 and 3.

The private treatment does not include the announcement, instead the experiment ends directly after Part 3 and subjects learn their payouts only from their computer screens. The crucial variation is therefore in the visibility of the dictators' decision: the public treatment effectively stimulates participants' image concerns. The

[3] To the best of our knowledge, such a randomization device has not been used in dictator games, but other studies have implemented comparable mechanisms to circumvent issues of self-selection, e.g. Dal Bó, Foster, and Putterman (2010) use a lottery with 90% and 10% probabilities to assign players to different games.

[4] Loewenstein (2000) explains how visceral factors such as fear are difficult to anticipate, while their actual appearance tremendously affects an individual's decisions making. We chose the direct response method, as our main motive of interest in DG 2 are a dictator's image concerns which basically coincide with a fear of stigmatization. By direct response we maximize the chances of directly suspending our subjects to this emotion. Brosig, Weimann, and Yang (2003) find a significant change in behavior for subjects under the direct response method when such non-monetary motives play a role.

payment is carried out privately in both treatments. Individual payoffs consist of the sum of points from the dictator game DG 1 in Part 1, the remainder of the auction endowment in Part 2, the dictator game DG 2 in Part 3, and a potential reward for one randomly chosen belief question.

1.3.3 Theoretical Considerations and Hypotheses

We draw our hypotheses from the incentive structure of a model based on Andreoni and Bernheim (2009) and Bénabou and Tirole (2006). To do so, we formally introduce objective functions for the dictator and recipient role. From there we proceed by backwards induction and derive our first hypothesis on the dictator offers in DG 2 in Part 3. Then we move to Part 2 and the second-price auction. We motivate the trade-off between the two roles and arrive at our second hypothesis on how the treatment variation affects bidding behavior.

Andreoni and Bernheim (2009) use a simplified version of the well-established fairness models and extend it by introducing image concerns as a further motive in an agent's utility function. A dictator's utility U_D increases in the expected type an external audience can infer from observing her action. For illustrative reasons we depart from the general utility function used by Andreoni and Bernheim (2009) to incorporate a more specific formulation, which builds on the seminal paper on image concerns by Bénabou and Tirole (2006).

Agent i in the role of the dictator D chooses what amount x_i of the surplus of 100 she will offer to her opponent j in the role of the recipient R. We assume her utility to be linear in consumption of $100 - x_i$. Her utility further depends on her intrinsic fairness type θ_i and the degree of image sensitivity μ_i, which are both private information. There is a commonly known continuous distribution function $F(\theta, \mu)$ over the entire domain $[0, \bar{\theta}] \times [0, \bar{\mu}]$ from which θ_i and μ_i are drawn. Both parameters $\bar{\mu}$ and $\bar{\theta}$ mark the highest manifestation of the respective types. The second term on the right hand side comprises her inequity aversion. The function $G(\cdot)$ is twice differentiable, strictly concave, and reaches a maximum of zero at zero. Any deviation of x_i from the fairness norm x^F[5] decreases the dictator's utility by factor θ_i, her own fairness preference. For higher θ_i the cost of unequal outcomes is greater, irregardless of who is advantaged. An individual with high intrinsic fairness concerns dislikes advantageous inequity in the same way as disadvantageous, while an individual with no fairness concerns does not care at all about the outcome distribution in this regard. The third term of the utility function comprises the image concern.

$$U_{iD}(x_i; \theta_i, \mu_i) = 100 - x_i + \theta_i G(x_i - x^F) - \mu_i(\bar{\theta} - E[\theta_i | x_i]) \tag{1.1}$$

[5]Technically, the parameter x^F can take on any value between 0 and 100 depending on the prevailing fairness norm. A commonly observed norm in bargaining games where roles are randomly assigned is the equal split ($x^F = 50$), which is a also assumed in Andreoni and Bernheim (2009).

While Andreoni and Bernheim (2009) and Bénabou and Tirole (2006) assume image to increase utility via the provision of an action, we argue that not providing an action – making a low offer – signals low intrinsic fairness concerns, which translate in an unfavorable social image and therefore come at a cost. Describing image concerns as costs is consistent with the empirical findings of Dana, Cain, and Dawes (2006). In their experiment, subjects in the dictator role frequently preferred to forgo additional surplus in order to escape their role and not be identified as dictators. We assume a dictator's image costs to result from a comparison between a benchmark fairness type $\bar{\theta}$ and an external audience's inference of her actual type based on her action $E[\theta_i|x_i]$. If the audience's inference coincides with the benchmark, the dictator's image is maximized and she experiences no costs from a bad image. If instead the inference falls short of the benchmark, the dictator experiences image costs in form of the difference between benchmark and inference multiplied by her own image sensitivity μ_i. Highly fairness concerned individuals, $\bar{\theta}$-types, who have no image concern ($\mu = 0$) serve as the benchmark. Their optimal offer $x^*(\theta = \bar{\theta}, \mu = 0) = x^*_{\bar{\theta}}$ presents the closest offer to x^F motivated by intrinsic fairness preferences alone and therefore portrays the purest behavior anyone will be measured against.[6] An external audience understands that any offer motivated by fairness alone is increasing in θ and must be in $[0, x^*_{\bar{\theta}}]$. By incorporating the external audience's inference into the dictator's objective function the individual choice problem becomes a signaling game. Image concerned individuals ($\mu > 0$) can exploit mixing with non-image concerned types and strategically choose their offer x_i to improve their image. Depending on the distribution $F(\theta, \mu)$, a given signal x_i can result in a conditional expectation $E[\theta_i|x_i]$ closer or further away from $\bar{\theta}$.[7]

The recipient R's utility U_R in DG 2 depends on the amount x_j she receives from the dictator j and her own fairness type θ_i. Her utility is linear in consumption of x_j and inequity aversion enters her utility in the same way as for the dictator.

$$U_{iR}(x_j, \theta_i) = x_j + \theta_i G(x_j - x^F) \tag{1.2}$$

In particular we assume that the recipient, contrary to the dictator, has no image concerns. She is passive in receiving the offer x_j and cannot act in any way. Consequently, an external audience has no signal to base a judgment on and therefore the recipient role presents an opportunity to avoid signaling one's type altogether. In the subsequent analysis of the auction, an agent will consider her utility from the dictator role U_D, which she actively affects by choosing x_i, against her utility from being the recipient U_R. In this ex ante consideration her utility as recipient depends

[6] Let μ_i be zero in (1) - the dictator has no image concern - then the optimal offer x^*_i for any $\theta_i > 0$ is never greater than x_F. The utility from consumption $(100 - x_i)$ is strictly decreasing in x_i and as $G(\cdot)$ is symmetric around zero, a deviation of x_i from x^F by a constant $c > 0$ causes the same utility loss due to fairness concerns for $x^1 = x^F - c$ as for $x^2 = x^F + c$. Since $x^1 < x^2$, utility from consumption is greater at x_1. Any optimal offer $x^*_i(\theta_i, \mu_i = 0)$ is never greater than x^F. Offers are also strictly increasing in fairness concerns θ_i, as costs from inequity increase.

[7] In this set-up even an offer $x_i = x^F$ can lead to image costs, depending on equilibrium strategies and the distribution $F(\cdot)$.

on what she expects another dictator j to offer $E_i[x_j] = x_i^e$. In the following we substitute x_j with x_i^e in the recipient's utility function.

A signaling equilibrium of this stage game consists of a dictator strategy $x_i^*(\theta_i, \mu_i) = \arg\max_{x_i} U_{iD}(x_i; \theta_i, \mu_i)$ for each type combination and consistent beliefs forming the conditional expectation $E[\theta_i|x_i]$ in (1). Solving for such equilibria is a complex task and beyond the scope of this text. Still, without formally determining all equilibria we focus on one specific kind and apply comparative statics to the optimal offer $x_i^*(\theta_i, \mu_i)$, which leads to our first hypothesis on the dictator offer in DG 2. In the private treatment, the dictator stays anonymous and she cannot be identified with her offer. We assume that image concerns either do not exist or only play a minor role in the private treatment. In the public treatment, the impact of image concerns intensifies. Formally expressed, the individual's image sensitivity μ_i increases in the public treatment and the commonly known distribution of μ shifts probability mass to higher values. In the following, we discuss how this affects the optimal offer x_i^*.

Differentiating the dictator's objective function (1) with respect to x_i and setting the derivative equal to zero gives the first-order condition for the optimal offer x_i^*.

$$F.O.C.: \quad \theta_i G_{x_i}(x_i - x^F) + \mu_i E_{x_i}[\theta_i|x_i] = 1 \tag{1.3}$$

For dictators without image concerns, $\mu_i = 0$, the first-order condition reduces to $\theta_i G_{x_i}(x_i - x^F) = 1$. Together with the previously made observation about purely fairness minded dictators, namely $x_i^* < x^F$, for each θ_i there exists exactly one offer x_i that balances the two sides of the equation. Therefore, for a given function $G(\cdot)$, there exists a strictly monotonic continuous function $x_i^*(\theta_i)$ that maps each θ-type to a corresponding optimal offer in the interval $(0, x^F)$.[8] An external audience understands this and, without image concerns present, would be perfectly able to infer the corresponding fairness type from an offer by the inverse function $x_i^{*-1}(\theta_i) = \theta_i(x_i^*)$. For dictators with image concerns $\mu_i > 0$, an optimal offer must balance the marginal benefit from consumption with the marginal costs from fairness and image concerns. It becomes clear that with the additional motive of image concerns, a given optimal offer $\bar{x_i^*}$ can result from multiple type combinations (θ_i, μ_i), which all balance both sides of the equation. Therefore, an external audience observing a given x_i cannot infer a respective fairness type with certainty. The variation in image sensitivity $\mu \geq 0$ enables image concerned dictators to pool with different fairness types by choosing the same x_i within $(0, x^F)$.[9] To see how image sensitivity affects image concerned dictators in their optimal behavior we apply the implicit function theorem on their first-order condition in (3). This informs us how their optimal behavior changes with their image sensitivity type.

[8] By construction, the interval is open and x^* only goes towards x^F in the limit as θ_i approaches greater values. The highest fairness type $\bar{\theta}$ makes the closest offer to x^F.

[9] Purely fairness concerned types will never choose an offer greater than x^F. Image concerned dictators with great image sensitivity might want to choose $x_i \geq x^F$, but we neglect this case as it seems to be of minor relevance empirically.

$$\frac{\partial x_i^*(\theta_i, \mu_i)}{\partial \mu_i} = -\frac{E_{x_i}[\theta_i|x_i]}{\theta_i G_{x_i x_i}(x_i - x^F) + \mu_i E_{x_i x_i}[\theta_i|x_i]} > 0 \tag{1.4}$$

By assumption the function $G(\cdot)$ is strictly concave and therefore the first term in the denominator is negative. The shape of the conditional expectation function $E[\theta_i|x_i]$ determines the sign of the entire term. In other words, the behavior of a dictator depends on how an external audience forms its expectation about her fairness type after observing an offer x_i. In the numerator, $E_{x_i}[\theta_i|x_i]$ is the marginal change in the expectation with respect to an increase in the offer. In the denominator, $E_{x_i x_i}[\theta_i|x_i]$ describes how the marginal change in the expected type changes with an increase in the offer x_i. We assume that a difference in offers is informative, so $E_{x_i}[\theta_i|x_i]$ cannot be zero over the entire interval $(0, x^F)$. For $E_{x_i}[\theta_i|x_i] > 0$ the expected type increases in the observed offer, while for $E_{x_i}[\theta_i|x_i] < 0$ the expected type decreases in the offer. Costs from an unfavorable image decrease in magnitude of the signaled fairness type. We assume that the information content of an offer decreases towards the more beneficial signal, as μ-types strive towards this direction. As more types make the same offer, the signal of this offer becomes less informative. Then in both cases marginal returns to an offer should not be increasing and $E_{x_i x_i}[\theta_i|x_i] \leq 0$. The entire denominator becomes negative and cancels out the negative sign in front of the fraction in (4). Finally, as image sensitivity μ_i increases, optimal offers x_i^* adjust in the same direction as conditional expectations on the fairness types increase.

A signaling equilibrium requires that beliefs inside the conditional expectation are consistent with behavior in equilibrium. From the first-order condition in (3) we see that if $E[\theta_i|x_i]$ is a strictly monotonic function, for a given optimal offer $\bar{x_i^*}$, the level curve of the optimal strategy $x^*(\theta_i, \mu_i)$ can be described by some strictly monotonic function between fairness and image types. Given $\bar{x_i^*}$ and θ_i, there exists exactly one μ_i which balances both sides of equation (3). It follows that strictly monotonic conditional expectations $E[\theta_i|x_i]$ are consistent with a strictly monotonic equilibrium strategy $x^*(\theta_i, \mu_i)$. From the discussion above about purely fairness minded dictators, we know that $x^*(\theta_i, \mu_i = 0)$ is a strictly monotonic increasing function in θ_i. As the distribution of types $F(\theta, \mu)$ places sufficient probability mass on purely fairness minded individuals, their equilibrium behavior presents a strong argument for the conditional expectation to be increasing in the observed offer ($E_{x_i}[\theta_i|x_i] > 0$). Image concerned dictators with low θ_i but high μ_i can take advantage of increasing conditional expectations in x_i. By choosing an offer x_i greater than the one they would choose only due to their own fairness considerations, they can pool with higher θ-types and decrease their image costs. The entire strategy $x^*(\theta_i, \mu_i)$ is monotonically increasing in both types. An increase in image sensitivity μ_i leads to an increase in the optimal offer x_i^*. Also with the fairness type θ_i in the denominator of (4), we find this effect to be stronger for low fairness types.

Our public treatment increases the visibility of the dictator offer x_i, thereby stimulates image concerns and increases the individual's image sensitivity μ_i. We should therefore observe rising offers in DG 2 under the public treatment.

Hypothesis 1
Dictator offers in DG 2 are higher in the public treatment compared to the private treatment.

Part 2 consists of the second-price auction followed by a lottery draw which assigns roles for DG 2 in Part 3. In the second price auction, two agents bid for winning the probability advantaged position on the dictator role in the lottery draw. The winner of the auction pays the second highest bid and becomes the dictator in the ensuing game with a 90% chance, while the loser only has a 10% chance. Each agent i has an endowment of 100 points for the bid $b_i \in [0, 100]$. The individual valuation V_i of winning the auction equals the expected utility from being in the probability advantaged position in the lottery draw minus the losing bid. V_i further depends on the respective types θ_i, μ_i, and the expected offer x_i^e when being in the recipient role. Losing the auction gives the expected utility based on the probability distribution from being in the disadvantaged position.

$$V_i(b_i, b_j; \theta_i, \mu_i, x_i^e) = \begin{cases} \frac{9}{10}\, U_{iD} + \frac{1}{10}\, U_{iR} - b_j & \text{, if } b_i > b_j \\ \frac{1}{10}\, U_{iD} + \frac{9}{10}\, U_{iR} & \text{, if } b_i < b_j \end{cases} \tag{1.5}$$

In equilibrium, both agents choose their strategies as best responses to each other's strategy. To identify the best response b_i for a given agent i, we determine a condition on the opponent's bid b_j under which the agent i prefers to win the auction. For a given type to be willing to win the auction, her valuation of winning needs to be greater or equal to her valuation of losing.

$$\frac{9}{10}\, U_{iD} + \frac{1}{10}\, U_{iR} - b_j \geq \frac{1}{10}\, U_{iD} + \frac{9}{10}\, U_{iR} \tag{1.6}$$

Rearranging (6) gives us an expression which states that an agent i is willing to win the auction if the anticipated difference in utilities between the two roles in DG 2 is greater or equal to a given bid b_j. To win the auction agent i would have to make a bid greater or equal b_j, but not greater than the left-hand side in (7), her maximum willingness to pay.

$$\frac{8}{10}\, (U_{iD}(x_i^*; \theta_i, \mu_i) - U_{iR}(x_i^e, \theta_i)) \geq b_j \tag{1.7}$$

For the opponent's bids b_j above her maximum willingness to pay, agent i prefers to lose the auction, as winning would lead to a loss. It can be shown that for any $b_j \in [0, 100]$ choosing $b_i(\theta_i, \mu_i, x_i^e) = \frac{8}{10}\, (U_{iD}(x_i^*; \theta_i, \mu_i) - U_{iR}(x_i^e, \theta_i))$ is a weakly dominant strategy. This leads to the well-known equilibrium of Vickrey auctions where both agents choose exactly their weakly dominant strategies $b_i^*(x_i^e, \theta_i, \mu_i)$. Moreover,

such an equilibrium has two convenient properties. First, it is strategy-proof, as strategies are not based on beliefs about the other agent's type and only appeal to an agent's own type. Second, the agents truthfully reveal their willingness to pay for the auctioned good. Here, the good is the probability advantage in the lottery for the dictator role. Because of the uncertainty induced by the lottery, we do not observe the willingness to pay for the dictator role per se, but instead an uncertainty-discounted equivalent. Primarily, we are interested in a change in the willingness to pay between our treatments. Observing a change in the discounted equivalent will therefore be just as informative.

In a next step we analyze how optimal bids b_i^* change with the treatment. We formally describe the public treatment as increasing an individual's image sensitivity. For an agent i this means that her optimal strategy is affected in two ways. First, her own image sensitivity μ_i increases, which directly affects the utility she anticipates from being in the dictator role U_{iD}. Second, she expects that her opponent's image sensitivity μ_j also increases, which indirectly affects the utility she anticipates from being in the receiver role U_{iR} through the expected offer x_i^e. To understand the effect of the treatment on the auction we first differentiate optimal bids b_i^* with respect to μ_i.

$$\frac{\partial b_i^*(x_i^e, \theta_i, \mu_i)}{\partial \mu_i} = 0.8 \frac{\partial U_{iD}(x_i^*(\cdot); \theta_i, \mu_i)}{\partial \mu_i} \tag{1.8}$$

By an increase in agent i's own image sensitivity μ_i, her anticipated utility from being in the dictator role U_{iD} changes and thus affects her equilibrium bid. To understand the direction of the change we take a closer look at how U_{iD} changes with μ_i in equilibrium. By applying the envelope theorem, the derivative turns out to be of the size of the image gap between highest possible fairness type and the conditional expected type.

$$\frac{\partial U_{iD}}{\partial \mu_i} = -(\bar{\theta} - E[\theta_i | x_i^*(\cdot)]) \tag{1.9}$$

For any offer the gap can never be negative and therefore the dictator's optimal utility decreases in image sensitivity μ_i.

Next we differentiate the optimal bid b_i^* with respect to μ_j.

$$\frac{\partial b_i^*(x_i^e, \theta_i, \mu_i)}{\partial \mu_j} = -\frac{8}{10} \frac{\partial U_{iR}(x_i^e, \theta_i)}{\partial \mu_j} \tag{1.10}$$

The opponent's image sensitivity μ_j enters into the recipient's utility U_{iR} through the expected offer x_i^e. The recipient herself has no image concerns, but her anticipated utility indirectly changes with μ_j in the following way.

$$\frac{\partial U_{iR}(x_i^e, \theta_i)}{\partial \mu_j} = \frac{\partial x_i^e}{\partial \mu_j} + \theta_i G_{x_i^e}(x_i^e - x^F) \frac{\partial x_i^e}{\partial \mu_j} \tag{1.11}$$

Her utility increases in the expected offer due to consumption and the decrease in inequity due to a higher offer. The expected offer in turn is affected by the public treatment in two ways.

$$x_i^e = E[x_j^*(\theta, \mu)] = \int_0^{\bar{\theta}} \int_0^{\bar{\mu}} x_j^*(\theta, \mu) f(\theta, \mu) \, d\mu \, d\theta \tag{1.12}$$

First, the receiver can expect the optimal dictator offer x_j^* to increase with μ_j, just as she anticipates for herself. Second, as the public treatment increases image sensitivity throughout the entire type space, probability mass in $F(\theta, \mu)$ shifts along the μ-dimension to higher values. The expected offer x_i^e increases with the public treatment. This in turn means that the recipient's utility U_R increases with the public treatment. An agent in the recipient role expects to get a higher offer from the dictator, which leads to more consumption and more fairness. Consequently, the recipient role becomes more attractive.

Combining the two effects implies how bidding behavior in the auction is affected by the public treatment. As the optimal bid $b_i(\theta_i, \mu_i, x_i^e) = \frac{8}{10} \left(U_{iD}(x_i^*; \theta_i, \mu_i) - U_{iR}(x_i^e; \theta_i) \right)$ constitutes the utility gap between the two roles, an increase in image sensitivity shrinks the gap by devaluing the dictator role and appreciating the recipient role. Therefore, auction bids decrease with the public treatment.

Hypothesis 2
Auction bids in DG 2 are lower in the public treatment compared to the private treatment.

1.4 Analysis

This section presents the analysis of the experimental data along the lines sketched in the previous theoretical discussion. To this end, we first introduce the key variables needed to test the hypotheses. Second, the protocol of the experimental sessions is briefly presented along with the characteristics of the subject pool. Afterwards, descriptive statistics and results of regression analyses are shown to test the hypotheses and in order to gain a more fine-grained understanding of the transmission channels between the public treatment and behavior. Following backwards induction, the analyses of the dictator game DG 2 are discussed before the auction.

1.4.1 Key Variables of Interest

To test the aforementioned hypotheses we generate the following variables from the experimental data: fairness, beliefs about others' fairness, auction bids, image concerns, and dictator offers in the public and private treatment (Table 1.1).

TABLE 1.1: Key Variables

	Min	Max	Mean	Std. Dev.
Fairness (DG 1)	0	31	12.4	8.9
Auction bid	0	100	49.0	35.1
Expected fairness/belief (DG 2)	0	100	30.3	19.4
Dictator offer (DG 2)	0	100	19.4	20.0
High image concern	0	1	0.32	0.47

In the initial dictator game DG 1 – played via the strategy method – subjects give on average 31 percent of the surplus to the recipients. The two most frequent shares are 20 points (equivalent to 50 percent of the surplus, chosen by 41 percent of subjects) and zero (chosen by 27 percent of subjects). DG 1 is used as a measure of the individual fairness preference θ in the absence of image concerns and role preferences.[10] There is no significant difference in mean offers in DG 1 between the two treatments ($p = 0.204$ t-test, $p = 0.139$ Mann-Whitney-U test (MWU)), showing that participants understood that this decision would never be made public and can therefore serve its purpose as a pure fairness measure. There are also no significant age or gender differences in the dictator offers. The variable *fairness* will comprise these dictator offers with a possible range from 0 to 40.

The auction bid is elicited from all subjects in the second-price auction preceding DG 2 and ranges from 0 to 100. Likewise, the offer made in DG 2 can range from 0 to 100, but is only observed for the subjects ending up in the role of the dictator.

In addition to subjects' actions, we elicit first-order beliefs through a standard incentivized mechanism in which subjects guess the actual behavior of others. As discussed in Section 3, the first-order belief about the dictator game DG 2 matters for the preceding auction. It is the amount x^e someone expects to receive from another dictator, in the further analysis named *expected offer*.

About one-third of subjects in the public treatment claimed (on a five-point Likert-scale) they were strongly or very strongly affected in their decision by the public announcement. For those subjects we construct the binary variable *image* sensitivity as a measure for μ, which is 1 for the subjects with the strong or very strong effect, and 0 for all others subjects (including all participants in the private treatment). As such it gives the subset of subjects in the public treatment who self-identify as especially image-concerned.

1.4.2 Protocol and Summary Statistics

The experiment was programmed using the software zTree (Fischbacher, 2007). Six sessions were conducted at the University of Hamburg in October 2020. 162 subjects

[10] Recall that the instructions for Part 2 and 3 were not handed out until after the end of Part 1. Before Part 1, participants were merely told that later on a task would follow in which their decision may or may not be made visible to others. The lab required this upfront announcement as the treatment slightly reduced the usually guaranteed degree of anonymity towards other participants. Subjects were allowed to reconsider whether they wanted to stay for the experiment. No subject chose to leave.

– recruited via the tool hroot (Bock, Baetge, and Nicklisch, 2014) – participated in one of the two treatments each (between-subjects design). 82 subjects participated in the public treatment and 80 in the control treatment. All sessions lasted for less than one hour.[11]

Upon arrival to the lab, subjects were placed anonymously in computer cubicles and received the instructions for Part 1 of the experiment. The instructions for Part 2 and 3 were handed out at the end of Part 1. After reading the instructions all subjects answered a set of control questions. At the end of each session, subjects completed a non-incentivized questionnaire on socio-demographic characteristics (e.g., age, gender, and field of study).Table 1.2 summarizes the most important subject pool characteristics.

TABLE 1.2: Summary Statistics

Total number of subjects	162			
Participants in private treatment	80			
Participants in public treatment	82			
	Min	Max	Mean	Std. Dev.
Age	19	46	26.5	4.9
Payout (in Euros)	5.1	14.3	9.1	1.9
Personal contacts per week	0	30	4.2	4.2
	Share			
Female	69%			
Economics/business student	27%			

1.4.3 Dictator Game (DG 2)

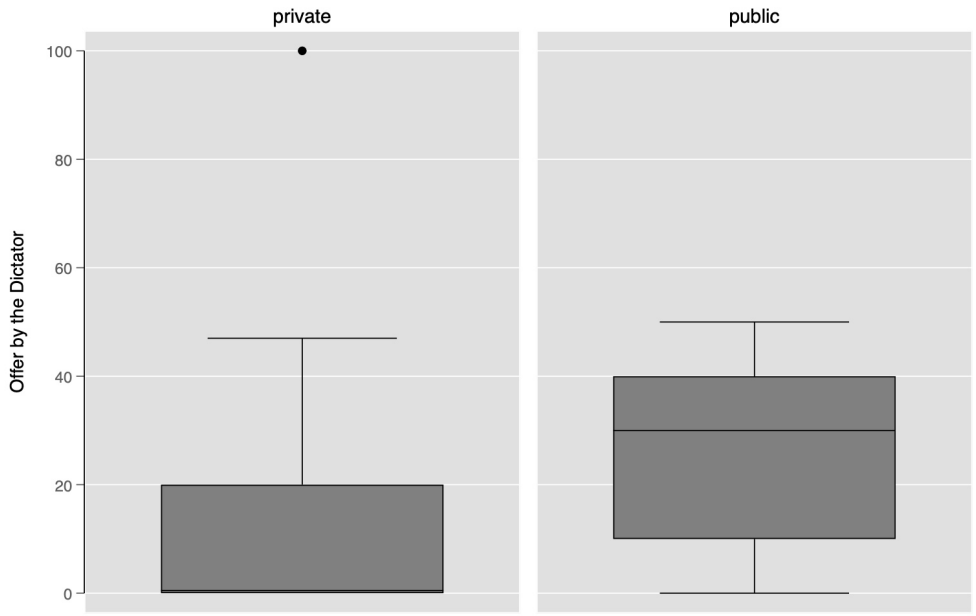
The dictator offers observed in DG 2 vary considerably between treatments (see Figure 1.2). The average dictator shares 14 points in the private and 24 in the public treatment ($p = 0.025$ t-test, $p = 0.005$ MWU). The distribution of offers is significantly different between the treatments ($p = 0.030$, combined Kolmogorov-Smirnov test). Exactly half of all dictators in the private treatment share nothing with their opponent, compared to only 22 percent in the public treatment. Vice versa, no subject in the private treatment shared 50 percent of the pie, compared to 15 percent in the public treatment ($p = 0.014$, chi-squared test). The observations are in line with the results from Andreoni and Bernheim (2009) and we cannot reject Hypothesis 1.

Result 1 *The distribution of dictator offers in the public treatment first-order stochastically dominates the distribution in the private treatment. Average dictator offers are higher and the equal split is more prevalent in the public treatment.*

To further investigate how the public treatment, fairness, and image concerns influence dictator behavior we continue with a multivariate analysis. The public treatment consistently has a positive and significant effect on the dictator offers in

[11] The experimental protocol ensured adherence to the local regulations against Covid-19.

FIGURE 1.2: Dictator Offers Part 3



regression analyses (Table 1.3). To account for the substantial number of zero offers we use left-censored Tobit regression models in addition to the OLS regressions.

All columns show not only a highly significant positive effect of the public treatment (binary variable *public*), but also regarding individual fairness. Column 1 shows that fairness increases offers in a one-to-one relationship. The evidence regarding an interaction effect between the treatment and fairness is ambiguous. The Tobit model suggests the treatment effect to vary for different levels of fairness. The treatment effect is largest when fairness equals zero. As an individual's fairness increases, the treatment effect diminishes. Highly fair-minded individuals are less affected by the stimulation of image concerns. This coincides with theoretical considerations about how the optimal offer changes with image sensitivity. Since fair individuals are intrinsically willing to offer more their offers are less affected by signaling motives.

Column 2 includes the variable image. In comparison to Column 1 the effect of fairness remains similar, while image sensitivity absorbs a large part of the treatment effect. A high image sensitivity increases average offers by 23 points. However, even the subjects who claim not to have been strongly affected by the treatment are significantly more generous with the public announcement.[12] The interaction between fairness and image is only significant in the Tobit regression in column four.

[12]One explanation suggests that subjects misreported their actual image sensitivity. This might be driven by image concerns towards the experimenter or simply by a lack of self-reflection. Another explanation confines the reason for higher offers in the public treatment to a selection process driven by lower bidding in the auction in Part 2. In section 1.4.4 we show that the average auction bid decreased in the treatment. If this change in bidding behavior between the treatments is driven by certain fairness types, then the sample of types becoming dictator potentially differs as well.

TABLE 1.3: Dictator Offers

Dependent Variable: Dictator Offer in DG 2				
	(1) OLS	(2) OLS	(3) Tobit	(4) Tobit
Public	15.15***	7.958*	36.86***	14.99***
	(2.783)	(3.270)	(8.839)	(5.029)
Fairness	1.084***	1.155***	2.229***	1.909***
	(0.170)	(0.131)	(0.344)	(0.254)
Public*Fairness	-0.305		-1.290***	
	(0.267)		(0.424)	
Image		23.48**		36.51***
		(5.949)		(7.780)
Image*Fairness		-1.262		-2.180**
		(0.776)		(0.975)
Constant	1.413	0.567	-24.91***	-19.49***
	(1.294)	(2.010)	(7.988)	(5.947)
N	81	81	81	81
R^2	0.254	0.333		
Pseudo R^2			0.057	0.071

Note: *Public* is a dummy variable for the treatment; *fairness* is the offer made in DG 1 (0 to 40); *image* is a dummy variable for self-reported treatment influence. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

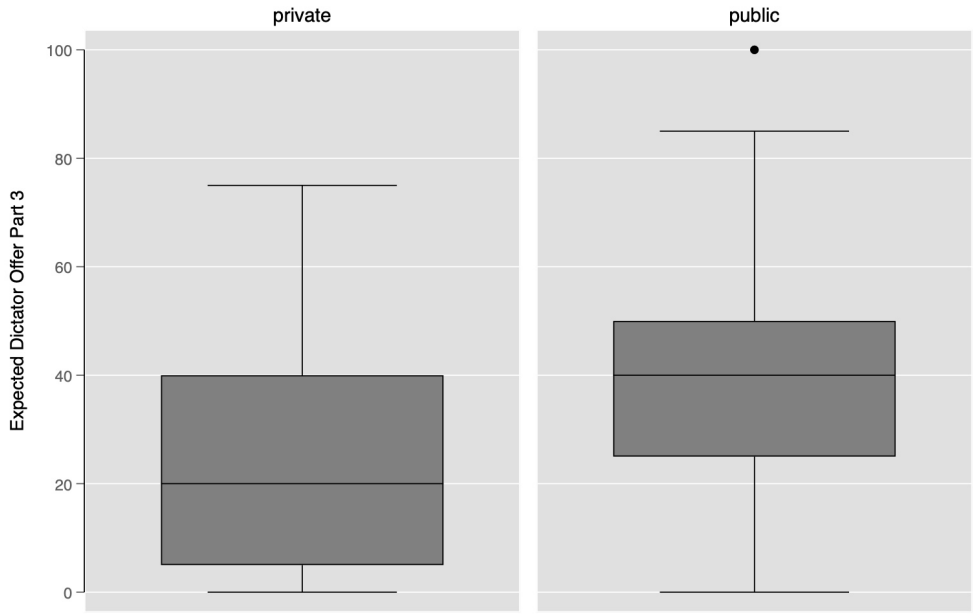
The treatment effect for individuals with high image sensitivity is greatest when they are selfish and decreases with fairness. The effect of fairness itself cancels out with high image sensitivity, as due to their image concern all such individuals are driven towards high offers anyways.

To summarize, the public treatment increases offers, which are in addition positively related to own fairness and image concerns. The initially observed heterogeneity in the treatment effect with respect to fairness becomes more distinct when interacting fairness with image, at least when correcting for censored data in a Tobit regression model.

After having grasped the effect of the public treatment on actual dictator behavior, we complete the investigation on DG 2 by analyzing how the treatment affects the expected dictator behavior x^e. Subjects expect on average 61% more points from the dictator in the public treatment compared to the private (37 vs. 23 points, $p < 0.001$ for t-test and MWU; see Figure 1.3). The distributions of expected offers are significantly different between treatments ($p < 0.001$, combined Kolmogorov-Smirnov test). We see that subjects respond to the treatment and correctly anticipate

observed behavior patterns. They anticipate others to not only be intrinsically prosocial but also image-concerned and therefore reacting to the public treatment.

FIGURE 1.3: Expected Offers in DG 2



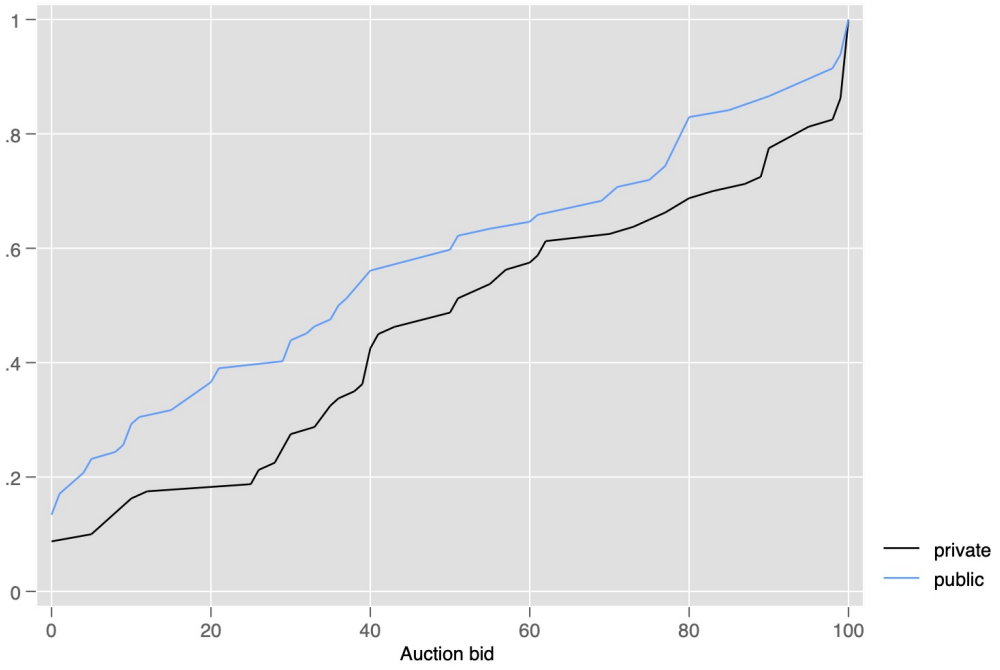
1.4.4 Bidding for the Dictator Role

In the pre-game auction in Part 2, the average bid amount is 43 points in the public and 55 in the private treatment (t-test $p = 0.035$, MWU $p = 0.024$). Bidding behavior reflects the stated preferences from the post-experiment questionnaire: significantly fewer participants in the public treatment prefer the dictator role over the recipient role (t-test $p = 0.012$, MWU $p = 0.019$). In the public treatment, the mode is bidding 0, whereas the mode in the private treatment is 100. Figure 1.4 shows the distribution of auction bids in both treatments. The distribution of bids in the public treatment first-order stochastically dominates the distribution in the private treatment. A Kolmogorov-Smirnov test rejects the hypothesis that both samples are drawn from the same distribution at a 5-percent significance level ($p = 0.038$ for the combined test). The evidence supports Hypothesis 2.

Result 2 *The distribution of auction bids in the private treatment first-order dominates the distribution in the public treatment. The average auction bid is lower in the public treatment.*

According to the theoretical considerations in Section 1.3.3, the optimal bid is a fraction of the difference between the anticipated utilities of dictator and recipient role. This difference and hence the bid vary with the visibility of the dictator's action – the public treatment –, more specifically it depends on the fairness preference, image sensitivity, and the expected dictator offer one receives as recipient. For the

FIGURE 1.4: Auction Bid



multivariate analysis we again use Tobit models as consistency checks for the OLS regressions (Table 1.4). The auction bids are limited from below at zero and above at one hundred. We find clustering at both values and therefore apply two-sided censored Tobit models. Adding interaction terms sheds light on how exactly the image concerns influence behavior.

Column 1 shows that fairness is significantly negatively correlated with the auction bid. An individual with higher fairness decreases her willingness to pay for the dictator role by around 0.8 points for every point sent in the dictator game in Part 1. There is neither a direct influence of the treatment itself, nor of the expected offer from other dictators on auction bids.

However, for the subset of image sensitive subjects – about one-third of the subjects in the public treatment – we see a remarkably low willingness to become the dictator, which becomes manifest in the large and negative effect of image on the auction bid (Columns 2 and 6). There is a significant heterogeneity in the treatment effect: for subjects without high image concerns, the public announcement has almost no impact on their role preference and the average bid of 51 points is statistically indistinguishable from the average bid of 55 points in the private treatment (t-test $p = 0.5513$). Whereas the average bid of image sensitive individuals is only 26 points (t-test $p < 0.001$).

TABLE 1.4: Auction Bids

	Dependent Variable: Auction Bid							
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Tobit	(6) Tobit	(7) Tobit	(8) Tobit
Public	-8.491				-10.03			
	(8.626)				(10.05)			
Fairness	-0.792**	-1.001***	-1.304***	-1.017***	-0.962***	-1.229***	-1.582***	-1.238***
	(0.206)	(0.158)	(0.219)	(0.166)	(0.317)	(0.285)	(0.326)	(0.289)
Expected Offer	-0.323	-0.247	-0.234	-0.324	-0.448*	-0.347	-0.330	-0.439
	(0.219)	(0.225)	(0.234)	(0.254)	(0.241)	(0.246)	(0.257)	(0.274)
Image		-28.62***	-45.63***	-63.80**		-35.49***	-55.99***	-78.57***
		(5.067)	(4.758)	(19.35)		(8.340)	(6.668)	(27.93)
Image*Fairness			1.790**				2.190***	
			(0.536)				(0.557)	
Image*Expected Offer				0.876				1.074*
				(0.456)				(0.649)
Constant	72.93***	73.50***	77.12***	75.88***	79.54***	80.31***	84.49***	83.02***
	(6.370)	(6.795)	(6.469)	(6.360)	(6.401)	(7.282)	(6.608)	(5.969)
N	162	162	162	162	162	162	162	162
R^2	0.114	0.180	0.208	0.198				
Pseudo R^2					0.024	0.028	0.026	

Note: *Public* is a dummy variable for the treatment; *fairness* is the offer made in Part 1 (0 to 40); *image* is a dummy variable for self-reported treatment influence; expected offer is the first-order belief about others' offers. Standard errors (clustered at session level) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This suggests that we cannot restrict attention to the average treatment effect itself, which would obliterate important differences between subjects. Once we control for image concerns directly, the negative influence of fairness becomes even larger and approaches a one-to-one relationship with the auction bid. The R^2 increases as well compared to Column 1, suggesting the model in Column 2 is an informative specification.

We further find that there is heterogeneity in the treatment effect with respect to fairness for the highly image sensitive individuals (Columns 3 and 7). The interaction effect between image sensitivity and fairness is positive, implying that selfish individuals are more strongly affected by the treatment. When focusing on the image sensitive subjects, the treatment lowers the average bid by 46 points for a completely selfish subject, but only by 10 points for a fully fair-minded one. It follows that much of the treatment effect on the role preference seems to stem from image sensitive individuals and especially those with low fairness concerns. The Tobit estimate in Column 7 corroborates this finding.

The correlation between expected offer and bid vanishes in the multivariate analysis. In Section 1.4.3 it was shown how the distribution of expected offer significantly shifted to higher values in the public treatment. We concluded that subjects anticipate others to be affected by the treatment. In contrast with our theoretical considerations, the expected sharing behavior of others does not matter for the auction bid once we control for the image concerns induced by the treatment. We therefore conclude that an agent's own image concern constitutes the first-order effect of the treatment on first-mover preferences.

Result 3 *Fairness and image sensitivity decrease auction bids. The treatment effect is driven by highly image sensitive individuals. For image sensitive individuals the treatment effect is heterogeneous with respect to fairness and decreases in fairness.*

1.5 Discussion and Conclusion

We study role preferences in relationship to image concerns in a dictator game experiment. Subject are matched in pairs and participate in a second-price auction. Winning the auction substantially increases the chances of becoming the dictator in the subsequent dictator game. The experimental design systematically varies the visibility of the dictators' sharing decision: in the private treatment, dictators' identities are never revealed to the recipients and no information about the amounts shared by other dictators is given. In the public treatment, at the end of the session participants gather in the laboratory hallway and all dictators' decisions are publicly announced.

Previous studies on making dictator decisions public showed ambiguous effects on dictator behavior (Hoffman et al., 1994; Bohnet and Frey, 1999; Dufwenberg and Muren, 2006; Andreoni and Bernheim, 2009). Not only do we shed more light on

this empirical controversy, but we are able to investigate an additional – previously omitted – aspect of the otherwise well-studied dictator game: how much subjects prefer the dictator role and how this preference changes with publicity. By applying the second-price auction we can quantify this role preference. While the dictator indisputably holds all the bargaining power in the interaction, she is also forced to reveal private information, e.g. about her fairness type, by making the sharing decision. We argue that this information revelation comes at a psychological cost and is therefore traded-off against the monetary gains made in the dictator position. Our theoretical argument is strongly inspired by Andreoni and Bernheim (2009)'s signaling model.

We find evidence that our treatment affected subjects in their decision-making at both stages of the experiment, in the auction and in the dictator game itself. In the dictator game, the public treatment increases the average offer by 10 points and the entire offer distribution is significantly shifted to the right. In a regression analysis we find that, indeed, someone's own fairness and image sensitivity matter for generosity in the dictator game. Next to the actual behavior of dictators we elicit subjects' first-order beliefs of their opponents' behavior as dictators. We find that with the public treatment the average expected offers increase by 15 currency units and the distribution significantly shifts to higher values. Subjects broadly seem to be convinced of an effect of the public treatment on their opponents' behavior. Finally, we find that the average auction bid significantly decreases by 12 points. The bid distributions significantly differ and shift to lower values in the treatment. By use of regression analysis we are able to identify own image concerns to have a first-order effect ahead of others' image concerns. We find a strong negative correlation between fairness, image sensitivity and auction bids. More specifically, the treatment effect is mainly driven by highly image sensitive individuals. We also find heterogeneity in the effect with respect to fairness. It turns out that the effect peaks for selfish individuals and decreases in fairness. We conclude that especially those selfish individuals would suffer image costs from publicity and therefore adjust their behavior to avoid informal punishment. Altogether this decreases their valuation of the dictator role and eventually reduces auction bids.

The relevance of the results beyond economic experiments is that increasing transparency has a significant influence on pro-social behavior. Players that care about how they are perceived by others are motivated to share more of a given surplus than they would do otherwise. They are also less likely to invest into gaining additional bargaining power. These insights matter for bargaining situations of very different kinds: wage negotiations (Rosenfeld and Denice, 2015), corporate social responsibility (Lee and Kohler, 2010), plea bargaining in criminal law (Schneider and Alkon, 2019), or international diplomacy (Stasavage, 2004) to name only a few. Most articles on the topic argue that transparency is beneficial because it reduces information asymmetries. Our results stress another aspect: transparency can reduce inequalities by directly changing the behavior of the more powerful bargaining party.

However, we show that the effect is strongest when actors care about how they are perceived. This can be the case, for example, if customers adjust their purchasing decisions in reaction to how a company is perceived by the public.

One aspect of the experimental design worthy of discussion is the fact that winning the auction only increases someone's chances in the following lottery to 90%. Undoubtedly, this complicates the understanding of the task and the calculation of the optimal bidding strategy for our subjects. Bidding on a lottery ticket inevitably invokes highly subjective considerations of risk preferences and perceptions of probabilities. Nevertheless, we prefer this procedure over an experiment without a lottery because we want to limit entitlement effects as much as possible. If it is commonly known that the subjects with the higher bids "bought" the dictator role with certainty, this changes the context of the dictator game substantially. The entitlement effect is well-researched and our design is therefore better suited for detecting insights about role preferences in the "regular" dictator game.

One aspect of the identification strategy that could be further developed in future studies on the topic is the measurement of image concerns. To create the measure of image sensitivity, we draw on the stated preferences how strongly affected subjects were by the public treatment. While there is no (monetary) incentive to misrepresent them, an objective measure would be more optimal here. Most studies which investigate the effect of social image concerns on behavior as a treatment make the assumption that all subjects are affected equally. Our evidence suggests that this is not necessarily the case. But we are not aware of an (economic) paper providing insights into reliably identifying subjects who can ex ante be expected to be more affected by such a treatment. An interdisciplinary approach seems promising here.

Bibliography

Andreoni, James and B Douglas Bernheim (2009). "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects". In: *Econometrica* 77.5, pp. 1607–1636.

Andreoni, James, Justin M Rao, and Hannah Trachtman (2017). "Avoiding the ask: A field experiment on altruism, empathy, and charitable giving". In: *Journal of Political Economy* 125.3, pp. 625–653.

Barmettler, Franziska, Ernst Fehr, and Christian Zehnder (2012). "Big experimenter is watching you! Anonymity and prosocial behavior in the laboratory". In: *Games and Economic Behavior* 75.1, pp. 17–34.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". In: *American Economic Review* 96.5, pp. 1652–1678.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Bohnet, Iris and Bruno S Frey (1999). "Social distance and other-regarding behavior in dictator games: Comment". In: *American Economic Review* 89.1, pp. 335–339.

Bolton, Gary E, Elena Katok, and Rami Zwick (1998). "Dictator game giving: Rules of fairness versus acts of kindness". In: *International Journal of Game Theory* 27.2, pp. 269–299.

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson (2007). "Is generosity involuntary?" In: *Economics Letters* 94.1, pp. 32–37.

Brosig, Jeannette, Joachim Weimann, and Chun-Lei Yang (2003). "The hot versus cold effect in a simple bargaining experiment". In: *Experimental Economics* 6.1, pp. 75–90.

Camerer, Colin F (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy". In: *American Economic Review* 100.5, pp. 2205–29.

Dana, Jason, Daylian M Cain, and Robyn M Dawes (2006). "What you don't know won't hurt me: Costly (but quiet) exit in dictator games". In: *Organizational Behavior and Human Decision Processes* 100.2, pp. 193–201.

Demiral, Elif E and Johanna Mollerstrom (2018). "The entitlement effect in the ultimatum game–does it even exist?" In: *Journal of Economic Behavior & Organization*.

Dreber, Anna, Tore Ellingsen, Magnus Johannesson, and David G Rand (2013). "Do people care about social context? Framing effects in dictator games". In: *Experimental Economics* 16.3, pp. 349–371.

Dufwenberg, Martin and Astri Muren (2006). "Generosity, anonymity, gender". In: *Journal of Economic Behavior & Organization* 61.1, pp. 42–49.

Engel, Christoph (2011). "Dictator games: A meta study". In: *Experimental Economics* 14.4, pp. 583–610.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Forsythe, Robert, Joel Horowitz, Nathan Savin, and Martin Sefton (1994). "Fairness in simple bargaining experiments". In: *Games and Economic Behavior* 6.3, pp. 347–369.

Frey, Bruno S and Iris Bohnet (1995). "Institutions affect fairness: Experimental investigations". In: *Journal of Institutional and Theoretical Economics*, pp. 286–303.

Frohlich, Norman, Joe Oppenheimer, and Anja Kurki (2004). "Modeling other-regarding preferences and an experimental test". In: *Public Choice* 119.1-2, pp. 91–117.

Frohlich, Norman, Joe Oppenheimer, and J Bernard Moore (2001). "Some doubts about measuring self-interest using dictator experiments: the costs of anonymity". In: *Journal of Economic Behavior & Organization* 46.3, pp. 271–290.

Gächter, Simon and Ernst Fehr (1999). "Collective action as a social exchange". In: *Journal of Economic Behavior & Organization* 39.4, pp. 341–369.

García-Gallego, Aurora, Nikolaos Georgantzís, and Ainhoa Jaramillo-Gutiérrez (2008). "Ultimatum salary bargaining with real effort". In: *Economics Letters* 98.1, pp. 78–83.

Güth, W and R Tietz (1986). "Auctioning ultimatum bargaining position". In: *Current Issues in West German Decision Research, Lang, Frankfurt, Germany*, pp. 60–73.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith (1994). "Preferences, property rights, and anonymity in bargaining games". In: *Games and Economic Behavior* 7.3, pp. 346–380.

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith (1996). "Social distance and other-regarding behavior in dictator games". In: *The American Economic Review* 86.3, pp. 653–660.

Johannesson, Magnus and Björn Persson (2000). "Non-reciprocal altruism in dictator games". In: *Economics Letters* 69.2, pp. 137–142.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1986). "Fairness and the assumptions of economics". In: *Journal of Business*, S285–S300.

Krupka, Erin L and Roberto A Weber (2013). "Identifying social norms using coordination games: Why does dictator game sharing vary?" In: *Journal of the European Economic Association* 11.3, pp. 495–524.

Laury, Susan K, James M Walker, and Arlington W Williams (1995). "Anonymity and the voluntary provision of public goods". In: *Journal of Economic Behavior & Organization* 27.3, pp. 365–380.

Lee, Matthew and Jillian Kohler (2010). "Benchmarking and transparency: Incentives for the pharmaceutical industry's corporate social responsibility". In: *Journal of Business Ethics* 95.4, pp. 641–658.

Loewenstein, George (2000). "Emotions in economic theory and economic behavior". In: *American Economic Review* 90.2, pp. 426–432.

Rankin, Frederick W (2006). "Requests and social distance in dictator games". In: *Journal of Economic Behavior & Organization* 60.1, pp. 27–36.

Rege, Mari and Kjetil Telle (2004). "The impact of social approval and framing on cooperation in public good situations". In: *Journal of Public Economics* 88.7-8, pp. 1625–1644.

Rosenfeld, Jake and Patrick Denice (2015). "The power of transparency: evidence from a British Workplace Survey". In: *American Sociological Review* 80.5, pp. 1045–1068.

Schneider, Andrea Kupfer and Cynthia Alkon (2019). "Bargaining in the dark: The need for transparency and data in plea bargaining". In: *New Criminal Law Review* 22.4, pp. 434–493.

Shachat, Jason and J Todd Swarthout (2013). "Auctioning the right to play ultimatum games and the impact on equilibrium selection". In: *Games* 4.4, pp. 738–753.

Stasavage, David (2004). "Open-door or closed-door? Transparency in domestic and international bargaining". In: *International Organization* 58.4, pp. 667–703.

A Appendix to Chapter 1

A.1 Instructions

Welcome! Today you will participate in an economic experiment. The duration of the experiment will be around 45 minutes. During this time we kindly ask you not to communicate with each other in any way. Questions should be directed only at members of the staff. Please raise your hand and we will be happy to come to your cubicle to answer any questions. Please also keep your cubicle's curtains closed. It is mandatory that you switch off your phones for the course of the experiment. Should you fail to comply with the laboratory's guidelines this will lead to you being excluded from the experiment without pay. At the end of the experiment, we kindly ask you to answer a short socio-economic questionnaire. The analysis of the data from this experiment will be completely anonymized such that none of your decisions can be linked to your personal data afterwards.

Please note that today there will be a slight deviation from the usual payment procedure. A part of your earnings from the experiment may be made visible to other participants of this session. It will be made clear which part of the experiment is concerned. If there are no further questions or concerns regarding the procedure we can now start with the experiment.

By participating in the experiment you will earn real money. How much you earn depends on your decisions and the decisions of other participants. Therefore, please read and follow the instructions carefully. Over the course of the experiment, you will earn or lose points. At the end of the session, your points will be converted into money by a ratio of 18 points for 1 €. The following experiment has **three parts**. The instructions for Part 2 and 3 will be handed out after Part 1. Please note that your choices in Part 1 does not influence the procedure and your potential earnings in the following parts in any way.

On several occasions today you will be asked to **make a guess about what the other participants are doing**. When asked to do so, please enter your personal assessment of the average behavior of the other participants in this session. At the end, one of those questions will be randomly picked and your answer is compared to the true observed behavior. Your earnings from this question will be determined in the following way:

You earn 18 points if your answer is in an interval of +/- 10 percent around the actual average.
You earn 9 points if your answer is in an interval of +/- 20 percent around the actual average.
You do not earn any points for this question if your answer is further away from the actual average. To summarize: The closer your guess is to the actual average, the more points will you earn for this question type.

Part 1

In Part 1, you will be randomly matched in pairs without knowing who your partner is. In these pairs you will play a game that has to roles: Player S (sender/she) and player R (recipient/he).

The game is as follows: **Player S receives 40 points and can distribute them freely between herself and Player R.** Player R only receives as many points as Player S is sending to him. Player S keeps the rest of the points for herself.

Your points from Part 1 are thus calculated as follows:
Points for Player S = 40 - points sent to Player E
Points for Player E = 0 + points sent from Player S

You will now have to decide how many points you want to send in case you become Player S. After every participant made this choice, the computer will randomly put together player pairs and again randomly decide who is Player S and who is Player E in each pair. Both participants in a pair are equally likely to become Player S or E. Your previous decision does not influence the matching of pairs or the role allocation. Player S's decision will be implemented in every pair.

You will learn at the end of today's session whether you were Player S or E and how many points you earned from Part 1. Please note that your earnings from Part 1 will never be made public.

Part 2 and 3

For Part 2 and 3 you will again be randomly put in pairs with another participant without knowing who he or she is. It is highly unlikely that you will be in a pair with the same person as in Part 1 and it is not possible to identify your partner.

In Part 3 you will again play a game with your partner which has to roles: Player A and Player B. In contrast to Part 1, it will be determined first who takes on which role. This is done in Part 2.

Part 2

Every player is now a part of a pair and receives 100 points. You can freely choose how many of these points (between 0 and 100) you want to use to increase your chances of becoming Player A in Part 3. Your partner simultaneously chooses an amount between 0 and 100 points. **The player who chose the higher amount of points will pay the amount of points which the other player has chosen.** The amount you pay is therefore never higher than the amount you chose. The player who chose the lower amount pays nothing and keeps the entire 100 points.

Your points from Part 2 are calculated as follows:
Player 1 chooses a higher amount of points than Player 2:
Points for Player 1 = 100 - chosen amount of Player 2
Points for Player 2 = 100

You will be informed about who chose more points in your pair at the end of the session. In the unlikely case where both of you chose exactly the same amount, the computer will break the tie and determine with equal chances who pays this amount.

Afterwards, every pair will take part in a lottery. The lottery is equivalent to blindly drawing a ball from an urn. There are nine blue balls and one red ball in the urn. If a blue ball is drawn, the player who chose the higher amount of points becomes Player A. If the red ball is drawn, the player who chose the lower amount becomes Player A. In both cases, the other player becomes Player B. The player who chose the higher amount therefore becomes Player A in 90 percent of the cases and the player who chose the lower amount in 10 percent of the cases. **Both players may become Player A, but it is more likely for the player who chose the higher amount of points.**

After the lottery every pair has a Player A and a Player B. The computer carries out the lottery in secret and will inform about the result, but not whether you or your partner chose more points.

Example 1:

- Player 1 chooses 47 points and Player 2 chooses 23 points.
- Player 1 chose the higher amount and pays 23 points (Player 2's amount). Player 2 pays nothing.

⇒ Player 1 keeps 77 points from Part 2 (100 - 23) and Player 2 keeps 100 points.

- In the lottery, the role of Player A will go to
  - Player 1 if a blue ball is drawn (9 out of 10 times)
  - Player 2 if the red ball is drawn (1 out of 10 times)

Example 2:

- Player 1 chooses 0 points and Player 2 chooses 89 points.
- Player 2 chose the higher amount and pays 0 points (Player 1's amount). Player 2 pays nothing as well.

⇒ Both players keep 100 points from Part 2.

- In the lottery, the role of Player A will go to
  - Player 2 if a blue ball is drawn (9 out of 10 times)
  - Player 1 if the red ball is drawn (1 out of 10 times)

Part 3

The player who is in the role of Player A after the lottery receives an additional 100 points. **Just as in Part 1, Player A can distribute these 100 points between herself and Player B.** Contrary to Part 1, you already know at this point whether you are the sending (A) or receiving (B) player. Player B receives the amount given by Player A, who keeps the remainder of the 100 points.

Your points from Part 3 are thus calculated as follows:
Points for Player A = 100 - points sent to Player B
Points for Player B = 0 + points sent from Player A

Payout

We kindly ask you to fill out an additional questionnaire. Your answer here do not influence your monetary payout. Afterwards, you will be informed of your earnings for today. Your payout will consist of the points you earned during the Parts 1,2, and 3 as well as the points from a randomly selected guessing question.

Points from Part 1
+ Points from Part 2
+ Points from Part 3
+ Points from a guessing question
= Total points
Total points ÷ 18 = Payout in €

Player B is informed on his computer screen how many points Player A has allocated to him and in which cubicle Player A is sitting. Afterwards, all participants

open the curtains of their cubicles. Furthermore, all participants who were in the role of player A stand up and step forward to the floor marking in front of their cubicle.

For each player A, the number of the cubicle is read out and it is announced how she has divided the 100 points in Part 3 between herself and player B. No further information, e.g., the name, will be made public. Player B will not stand up and Player A will not be able to identify her partner. Every participant today will thus be informed about every Player A's sharing decision.

Afterwards, we will call you one by one to receive the full payout. Only you and one lab assistant will know the full amount of your earnings today.

Chapter 2

The Influence of Democracy and Leadership on Cooperation

Author Fanny Schories

Abstract The paper examines whether an institution has a differing impact on cooperation if it is introduced by a representative of the affected subjects rather than exogenously imposed. The experimental design controls for selection effects arising from the endogenous policy choice. The treatment varies whether the decision-maker is elected or randomly appointed. There is evidence of a large democracy premium in the sense that endogenously chosen institutions lead to more cooperation than identical exogenous institutions, but only if the group leader is democratically chosen. Especially the subjects who initially did not prefer the policy are more likely to cooperate if it was brought about by an elected representative. There is no such democracy premium for randomly appointed group leaders.

Keywords Laboratory Experiment, Representative Democracy, Collective Decision-Making, Social Dilemma, Legitimacy.

JEL Classification C9, D02, D72.

2.1 Introduction

Does the way a law is implemented influence its effectiveness? An experiment is used to quantify the behavioral difference between exogenously formed institutions and those that are implemented through a democratic procedure. This behavioral difference has been coined *democracy premium*: All else equal, there is an increased willingness to cooperate after an institution was introduced through a democratic procedure.

That a central authority can effectively improve cooperation is a stylized fact in experimental economics. However, most studies take these institutions as exogenously given. Recently, many experiments studied endogenous institutions which do not "fall from heaven" but are introduced by the affected parties themselves. The process that leads to an institutional setting might well influence how effectively it can fulfill its societal purpose. Should a law that was effective in one instance also be assigned in other situations? Consider the progressing European integration as an example. Is a reform introduced in Greece as effective when it is de facto prescribed from an external authority such as the "Troika" as if it was introduced autonomously by the elected national government?

Previous studies found inconclusive evidence regarding the existence of a democracy premium in direct democracies (Dal Bó, Foster, and Putterman, 2010; Sutter, Haigner, and Kocher, 2010; Vollan et al., 2017; Gallier, 2020). This study focuses on representation as another aspect of democratic decision-making. Indirect democratic procedures are commonly used: nations, firms, and clubs typically delegate at least parts of their decision-making processes to representatives. The paper contributes to two central questions: Is there a democracy premium when a representative chooses the relevant institution? And does the democracy premium depend on how the representative came into office?

The two experimental treatments address the second question. The treatment variable is the way the group leader is appointed: via election or lottery. The treatments are therefore called indirect democracy (ID) and random dictator (RD). Indirect – or representative – democracy is the most common form of democracy today (Alizada et al., 2021). Using lots to assign political roles, also known as *sortition*, dates back to Athenian democracy, where it was promoted by Aristotle to achieve the democratic ideals of equality and fairness better than elections (Barnes, 1984).

The experiment has three stages. The first stage of the experiment consists of a prisoners' dilemma played in small groups. In the second stage, subjects form preferences about a payoff modification for their group in the final stage. The modification transforms the prisoners' dilemma into a coordination game, which makes both defection as well as cooperation incentive-compatible. In the indirect democracy treatment, subjects elect a group representative, whereas a group leader is determined by chance in the random dictator treatment. The representative or leader's preference about changing the payoffs becomes binding for the group but is only

considered in 50 percent of the cases. If it is not considered, either the coordination game or prisoners' dilemma is randomly assigned to each group for the third stage.

This design feature – first introduced by Dal Bó, Foster, and Putterman (2010) (hereafter: DFP) – can control for information and selection effects and thus allows a clean estimate of the democracy premium. Cooperation rates before and after the vote are analyzed conditional on individual policy preferences and the outcome of the random intervention to estimate the democracy premium for both treatments separately.

The payoff modification significantly fosters cooperation and even more so when introduced by an elected representative. Moreover, the difference between treatments is striking. There is a substantial effect of endogenous choice in the representative democracy: 78 percent of the total policy effect can neither be attributed to the payoff change itself nor to differences in group composition and thus remain as the democracy premium. In contrast to DFP (2010), especially those who initially do not prefer the coordination game respond strongly to a democratic payoff modification with an increased willingness to cooperate. Conversely, the randomly appointed leader does not bring about an increase in cooperation beyond the exogenous payoff modification. The only effect of the random dictator is negative because cooperation is further decreased in the case of endogenous non-modification. The results have important implications for policy-making but also for the methods used in experimental economics.

The remainder of the paper is structured as follows. Section 2 reviews the related literature focusing on the effects of endogenous formal institutions in economic laboratory experiments. Section 3 presents the design of the experiment, including testable hypotheses. The analysis and results are presented in Section 4. Finally, Section 5 discusses potential explanatory approaches from economic theory and concludes.

2.2 Related Literature

The focus of this paper lies on formal institutions that are exogenously enforced in the form of law or other regulation, as opposed to informal sanctions, which are maintained privately. Elinor Ostrom laid the groundwork for the experimental study of self-governance as a way of overcoming collective action problems (e.g. Ostrom, 1991; Ostrom, Walker, and Gardner, 1992). By now, there is a large and growing body of experimental literature exploring the key factors that influence cooperative behavior in societies. These studies suggest that the implementation of an institution matters in addition to the institutional design itself. A central result

from previous lab experiments on such institutions is that direct democratic participation rights increase subjects' contributions to a public good, ceteris paribus.[1] These insights cannot be explained with outcome-oriented utility concepts as used in, for example, rational choice theory (Becker, 1976). A cooperation premium of democratic institutions should not exist under rational choice if the institution that is implemented and the information provided remain the same.

The evidence about the influence of participation rights on cooperation levels is based mainly on public goods experiments, where democratic structures are implemented into the policy selection process by allowing participants to vote on different proposals directly (see Dal Bó (2014) and Dannenberg and Gallier (2020) for surveys of the literature on endogenous institutions). Tyran and Feld (2006) show that an endogenously chosen non-deterrent law reduces free-riding behavior. The experiment varies the severity and enaction of a monetary punishment on free-riding. An exogenously implemented mild law does not significantly increase compliance compared to the game without law. In the endogenous treatment, individuals mostly vote in favor of the mild law, and the contribution rate is significantly higher than without law (Tyran and Feld, 2006). Sutter, Haigner, and Kocher (2010) present additional evidence that participation rights reinforce cooperation. Subjects vote for a decentral punishment or reward mechanism. The endogenous choice is associated with higher contributions for any given institution compared to an identical mechanism implemented through an external authority (Sutter, Haigner, and Kocher, 2010). Markussen, Putterman, and Tyran (2014), Kamei, Putterman, and Tyran (2015), and Dannenberg, Haita-Falah, and Zitzelsberger (2020) find that endogenously chosen sanctioning mechanisms improve public good provision. However, the experiments presented so far in this section cannot isolate a pure democracy premium. The vote entails a signaling component and reveals information about the group composition and subjects' preferences in the endogenous case. Conditionally cooperative players are likely to respond to this signal and adjust their behavior accordingly. Furthermore, because of the democratic policy selection, the institution is not randomly assigned, and the estimated differences between exogenous and endogenous assignment are potentially biased by self-selection. Both self-selection and information as confounding factors are mitigated using the experimental design presented in Section 2.3.1.

The experimental mechanism employed in the present study was first introduced by DFP (2010). It avoids a self-selection bias in experiments investigating the effect

[1] Various authors have taken the search for effects of endogenous institutions in public goods games to the field. Cavalcanti, Schläpfer, and Schmid (2010) find that public deliberation increases the willingness to contribute to projects for the management of common resources among Brazilian fishermen. Other studies such as Bonin, Jones, and Putterman (1993), Bardhan (2000), Black and Lynch (2001), and Fearon, Humphreys, and Weinstein (2011) find similar results in settings ranging from irrigation rules in rural India to workplace decisions of manufacturing businesses in the USA: participation rights increase compliance, productivity, and satisfaction. Grossman and Baldassarri (2012) find that subjects electing leaders contribute more to public goods than subjects who were assigned leaders through a lottery.

of democracy. Subjects are allowed to vote on a fixed policy proposal, but this democratic choice is overruled by a random computer decision in 50 percent of the cases. This strategy makes it possible to control for unobservable characteristics that influence both voting decisions and cooperative behavior. DFP (2010) find that even when controlling for selection, endogenous and exogenous institutions have a differing impact on cooperation. The authors find evidence of a democracy premium: a cooperation-enhancing influence of democratic institutions beyond the instrumental effect of the policy choice. Sutter, Haigner, and Kocher (2010) obtain contradictory results using a similar randomization mechanism: whether the vote was considered has no significant influence. The authors conclude that the institutional design itself influences behavior and not its democratic implementation. Kamei (2016) follows the randomization mechanism suggested by DFP (2010) and finds not only a direct democracy premium but also pro-social behavioral spillovers from democratic procedures: Those involved in an endogenous institution formation keep their increased cooperativeness even in a following setting without democracy. Gallier (2020) combines the experimental techniques used in Tyran and Feld (2006) and DFP (2010). He finds ambiguous results regarding the democracy premium: overall, contributions to a public good are higher if a sanctioning institution is democratically implemented. However, the difference is driven mainly by self-selection and information effects, which the identification strategy can precisely estimate. Only for the subjects that initially did not prefer the institution does an actual democracy premium exist. Vollan et al. (2017) find the opposite of a democracy premium in China. Instead, the Chinese subjects are more prone to comply with an exogenous policy. The authors attribute this result to a culture of obedience towards authority. No positive democracy premium has been found in a representative democracy yet (Castillo et al., 2017; Kamei, 2017).

The treatment variation presented in this paper directly connects to the strand of literature on the differences between elected and appointed leaders. A regularity already established in the literature is that elected leaders behave more pro-socially than randomly appointed ones for several reasons, e.g., because the election functions as a screening tool, because it creates accountability or because the leaders demonstrate reciprocity towards voters (Brandts, Güth, and Stiehler, 2006; Hamman, Weber, and Woon, 2011; Corazzini et al., 2014; Brandts, Cooper, and Weber, 2015; Marcin, Robalo, and Tausch, 2019; Drazen and Ozbay, 2019). The threat of reelection is not present in the paper here, such that it is most closely related to Drazen and Ozbay (2019), who compare the behavior of elected and non-elected leaders in a citizen-candidate model setting and find that only the elected representatives demonstrate reciprocity towards the citizen. What differentiates this paper from the previously cited studies is that here the leader takes only one decision on behalf of the other subjects but then continues to act as a regular group member. Furthermore, voters have been found to be more satisfied and sustain higher cooperation if they were involved in selecting a leader (Rivas and Sutter, 2011; Andreoni and

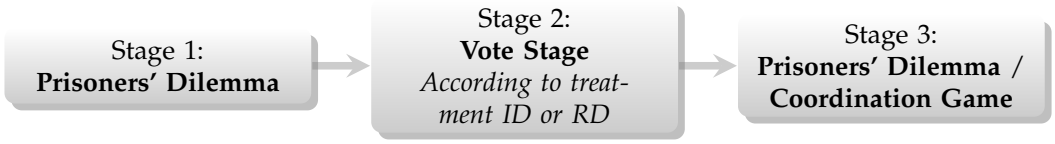
Gee, 2012; Mechtenberg and Tyran, 2019). Section 2.3.3 returns to the findings from the literature presented in this section and how they inform the hypotheses for the experiment conducted in this study.

2.3 The Experiment

2.3.1 Experimental Design

The aim of the experiment is (1) to investigate whether a policy that a group leader actively chose induces more cooperation than the same policy implemented via an exogenous mechanism, and (2) whether in addition the way the leader making the endogenous choice comes into office influences subjects' willingness to cooperate. Both questions are answered using a between-subjects design. (1) is investigated using the identification strategy of DFP (2010). For the second part of the research question, the two treatments vary how the leader is chosen: through an indirect democratic process [ID treatment] or randomly appointed [RD treatment].[2] The experiment consists of three stages (see Figure 2.1).

FIGURE 2.1: Sequence of the Experiment



The games are based on a standard prisoners' dilemma, in which players have the choice between two actions: cooperate and defect (Table 2.1).[3] The prisoners' dilemma has a unique symmetric Nash equilibrium where both players defect. Following DFP (2010), ten rounds of the prisoners' dilemma are played in the first stage with random rematching of pairs in every round within each group. The groups are made up of four players and remain together over the entire session. When deciding on an action in the prisoners' dilemma, players do not know with whom they are paired. But after each round, the screen informs them who their opponent was and what action each group member chose.

		Player 2	
		A	*B*
Player 1	*A*	(50, 50)	(30, 60)
	B	(60, 30)	(40, 40)

		Player 2	
		A	*B*
	A	(50, 50)	(30, 48)
	B	(48, 30)	(40, 40)

TABLE 2.1: Prisoners' Dilemma (left) and Coordination Game (right)

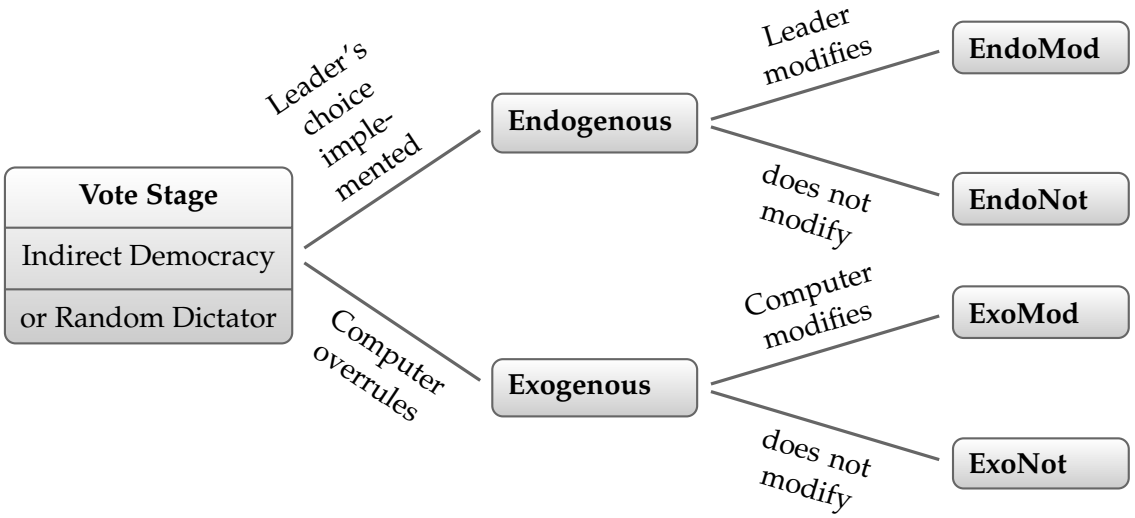
[2] Instructions are included in the appendix. A replication attempt of DFP (2010)'s direct democracy [DD treatment] can also be found in the appendix.

[3] To preserve neutral framing, the actions are labeled A (cooperate) and B (defect) in the experiment's instructions.

In the following vote stage, subjects make one or two choices depending on the treatment. In both treatments, subjects first decide whether they want to change the payoff of their group to a coordination game for stage 3 or remain with their group in the prisoners' dilemma. The coordination game has a Pareto-superior Nash equilibrium in (A,A) (mutual cooperation). However, (B,B) (mutual defection) remains as a Nash equilibrium also in the coordination game. Every subject privately states a preference whether to implement the payoff modification or not. This decision will matter if she becomes the group leader. Second, the leader is determined according to the treatment. In the democracy treatment (ID), to elect the representative players privately announce another group member's player identification number without knowing her preference for modification. The information available for this election is that there is complete information about the actions of every group member in the first ten rounds of the experiment. The player whose identification number is named most often in a group is elected as representative with plurality rule. Any tie is broken by the computer. In the random dictator treatment (RD), players do not vote for the representative. Instead, one player from each group is selected by a lottery, placing the same probability on every subject. Now every small group has a leader, either randomly appointed or democratically elected, and this leader has stated a game choice for stage 3, which becomes binding for the group.

Analogous to DFP (2010), each leader's preferred game is implemented with a 50 percent probability. If it is not implemented, the computer chooses either the prisoners' dilemma or coordination game for the group, again with a 50 percent probability for each game. Consequently, there are four conditions under which subjects play stage 3 (see Figure 2.2): payoffs modified to a coordination game by the representative (EndoMod) or by the computer (ExoMod), and the unmodified prisoners' dilemma game either chosen exogenously (ExoNot) or by the leader (EndoNot).

FIGURE 2.2: Four Possible Vote Stage Outcomes (adapted from DFP, 2010)



The twofold random intervention makes it possible to compare groups whose leaders decided in the same way but ended up in different conditions. The intervention is crucial to control for self-selection. Assuming there are unobservable player characteristics that increase both the preference for the coordination game as well as the willingness to cooperate, the treatment assignment is non-random whenever subjects choose their own payoff structure (by voting or otherwise). However, as DFP (2010) show, once the analysis conditions on an individual's original preference and the payoff structure, correlated unobservable characteristics are controlled for.

Immediately after the vote stage, subjects are informed about the leader's player ID and game choice, whether the choice was considered, and the game the group will ultimately play in the last stage. Consequently, the experimental design controls for payoff modification preferences via the strategy method and holds information constant across outcomes (see discussion below). The respective game is then played for another ten rounds in Stage 3.

2.3.2 Self-Selection and Information Effects

Two aspects of the design are worthy of discussion, as they are inherently relevant to all experiments on endogenous institutions: self-selection and information effects. The main confounding factor of experiments investigating democracy is self-selection: whenever a policy is endogenously introduced, participants select into the treatment by definition. It is natural that players vote for the policy that is aligned with their preferences, making it difficult to compare the impact of different institutional designs since the assignment is not random. Cooperative subjects are more likely to prefer the policy that fosters cooperation.

Self-selection leads to an overestimation of the effect of endogenous policy selection since the voting decision and behavior are positively correlated (DFP, 2010). The randomization and control for voting behavior are introduced into the experiment to eliminate the self-selection effect from the analysis. The individual vote for or against the payoff modification can serve as a control for unobserved underlying characteristics influencing the willingness to cooperate. This design holds the advantage that the results of subjects that voted in the same way and ended up with the same game, but through a different mechanism, can be compared. DFP (2010)'s identification strategy relies on the assumption that groups with an identical distribution of votes for and against modification also have identical preferences about modification and thus cooperation. If their behavior differs, this is attributed to the way the modification was implemented.

The second confounding factor that the design controls for is information. Sutter, Haigner, and Kocher (2010)'s as well as DFP (2010)'s main experiment involve instructing subjects about their group's choice only in the endogenous case, thereby

straining the ceteris paribus assumption: not only does the institution's implementation differ between the endogenous and exogenous condition, but also the information provided. One can argue that this information is an essential part of a democratic institution, and the asymmetry between the outcomes should not be erased. But in this case, it is not possible to truly isolate the democracy premium from information effects. Gallier (2020) systematically varies the informational content in the exogenous condition and finds that the information effect is a driver of heterogeneities in response to the policy implementation. In the present experiment, subjects are thus informed about the leader's choice in the endogenous *and* exogenous outcomes. The design ensures that there are no differences in the amount or quality of available information between the endogenous and exogenous vote stage outcomes that subjects could condition their behavior on in the following rounds, except with regard to the policy implementation according to the treatment. If the leader's choice of game was revealed only in the endogenous condition, subjects would be able to update their beliefs about their group members in a way that the subjects in the exogenous case could not.

Holding the available information constant across outcomes means that observed differences can be attributed to the intrinsic difference between democratically and exogenously introduced policies. I consider this a conservative estimate of the democracy premium because both above-mentioned related factors inherent to democratic processes are controlled for.

2.3.3 Hypotheses

Following previous literature and the discussion in this section so far, we can now establish specific hypotheses to be tested by the experiment. Unless explicitly stated, all hypotheses apply to both treatments and are tested separately and between subjects. It is a weakly dominant strategy for subjects in both treatments to choose the game they truly prefer, a necessary assumption for the identification strategy. The individual game choice never influences who becomes the group leader: the game preference of a subject remains private information during the election in ID, and leaders are chosen randomly in RD. Hence, there is no incentive to misrepresent game preferences for the sake of becoming group leader, and the game choice is strategy-proof in that regard. Furthermore, the optimal choice between the games does not depend on the treatment.

Which game is preferred depends on a subject's beliefs about the other group members' actions in the coordination game. If subjects expect to achieve mutual cooperation under the modification, they should vote in its favor. whereas they are indifferent between the two games if they expect mutual defection in the coordination game (DFP, 2010). In the latter case the equilibrium of (D,D) will be played in any case, leaving no incentive to introduce the modification. Off-equilibrium reasoning can bring subjects to prefer the prisoners' dilemma if they expect to defect against a cooperating player: the payoff modification decreases the deviation payoff in this

case from 60 to 48 points. As the payoff modification makes the Pareto-superior outcome of mutual cooperation an equilibrium we can expect both unconditionally as well as conditionally cooperative players to strictly prefer modification. Uncooperative subjects, on the other hand, have a weak preference for the prisoners' dilemma.

Hypothesis 3 *A subject who is cooperative in stage 1 is more likely to have a preference for payoff modification than an uncooperative subject.*

The experiment is able to test three factors influencing cooperative behavior that are identified in the literature on the effects of elected leadership: selection of more cooperative leaders, intrinsic motivational changes of elected leaders', and changes in voter behavior in response to the decision-making process (Drazen and Ozbay, 2019). While the emphasis is on the latter, which includes the democracy premium phenomenon, the former two aspects are also testable using the experimental design at hand.

First, according to the selection factor, it is to be expected that groups in ID elect the most cooperative subject as representative (Hamman, Weber, and Woon, 2011). It follows directly from Hypothesis 1 that the highly cooperative subjects should vote for another cooperative subject. Subjects who are conditional cooperators and play equilibrium strategies in which they defect in the prisoners' dilemma but cooperate in the coordination game are better off under the modified payoffs and should thus vote for a more cooperative subject as well. Only subjects who always defect can have a weak preference for another non-cooperative subject as representative under the assumption that she will not modify payoffs and there are some other subjects in the group who cooperate off the equilibrium path. In short, only off-equilibrium beliefs justify having strong preferences for either the unmodified payoffs or an uncooperative representative.

Hypothesis 4 *In treatment ID, cooperative players are more likely to be elected as representatives.*

Second, elected leaders have been shown to act less selfishly than randomly chosen leaders (Drazen and Ozbay, 2019). This can be tested via a comparison between the two treatments.

Hypothesis 5 *In Stage 3, leaders in ID behave more cooperatively than leaders in RD.*

Third, the response of the voters to the implementation process is at the heart of the paper. Rational choice would predict subjects to be indifferent between the decision-making procedures ID, RD, and exogenous as long as the outcome, including the information revealed about other players' preferences, remains constant. However, as discussed in Section 2.2, numerous empirical studies have rejected this null hypothesis for direct democratic decision-making. We could therefore expect

endogenous policy selection to lead to more cooperation than the exogenously imposed policy in both treatments (between subjects comparison in Stage 3). Regarding the difference between direct and representative democracy, the direct process seems to be seen as carrying higher legitimacy (Olken, 2010; Towfigh et al., 2016). Therefore, we can expect the democracy premium to be smaller in the present setting compared to DFP (2010).

Hypothesis 6 *There is a democracy premium, i.e. cooperation rates are higher when a policy is democratically introduced.*

Finally, the treatments ID and RD serve to improve the understanding of the sources of the democracy premium. Two candidate transmission channels shaping compliance with the outcome of a decision-making procedure are legitimacy and authority. On the one hand, legitimacy can be narrowly defined as being derived from the consent of the governed through elections (Locke, 1983). More specifically, institutions which are responsive to citizens by allowing participation in the decision-making process carry *input legitimacy* (Schmidt, 2013). On the other hand, authority is concentrated political power, regardless of its source. If the two coincide, a feeling of mutual responsibility arises. This is the case in a representative democracy, where authority stems from the figure of the representative herself and legitimacy from her democratic election. A random dictator has authority but no legitimacy from the political process. Compliance is described as a crucial consequence of political legitimacy (Ham et al., 2017). Thus, depending on which treatment corresponds to higher cooperation rates following the decision-making process, we can isolate the more important transmission channel of the democracy premium between the two factors authority and legitimacy, leading to two competing hypotheses.

Hypothesis 7.1 *(Legitimacy) The democracy premium is larger in ID than RD*

Hypothesis 7.2 *(Authority) There is no difference between the democracy premia in ID and RD.*

2.4 Analysis

2.4.1 Protocol and Summary Statistics

10 sessions took place at Hamburg University between 2016 and 2019 with a total of 280 participants (140 per treatment). No subject participated in more than one session or treatment. Upon arrival to the lab, subjects were randomly assigned to a computer cubicle and received the instructions in written form.[4] At the end of

[4] A translation of the originally German instructions can be found in the appendix. Instructions were read aloud and every subject correctly answered a set of control questions to ensure the instructions were well understood. Instructions for the second and third stage were handed out after the end of the first stage in order not to influence behavior prior to the vote.

the session, participants filled out an unpaid socio-economic questionnaire including three questions from a cognitive reflection test to elicit strategic sophistication (Frederick, 2005).

Table 2.2 presents summary statistics. 106 subjects identified themselves as male, 166 as female.[5] The laboratory uses the software hroot (Bock, Baetge, and Nicklisch, 2014) to invite subjects from a pool of around 7,000 registered participants recruited on the main campus at Hamburg University. Hence, the vast majority of subjects are full-time students. Out of those, around one-third of students were economics or business majors. More than a quarter of the subjects stated that they had at some point taken a class in game theory. More than one third of the subjects answered all three of the aforementioned logic questions correctly.[6] Payment was made according to the outcome of two randomly chosen rounds from the first and third stage respectively with an exchange rate of 10 points = €1. Subjects earned €9 on average which is in line with the mean hourly wage of €10 that the lab promises since all sessions lasted less than an hour. Subjects privately collected their payment in cash at the end of a session.

TABLE 2.2: Summary Statistics

	Min	Max	Mean	Std. Deviation
Age	17	57	25.7	5.1
Correct Logic Questions	0	3	1.7	1.2
Payout (€)	6	12	8.8	1.2
	Share			
Female	59 %			
Full-time Student	97 %			
Economics Student	31 %			
Game Theory Knowledge	29 %			

Note: Sample size is n = 280 with 140 subjects in each treatment.

2.4.2 Individual Analysis

The following sections examine the experimental data in the light of the hypotheses. Unless indicated otherwise, *p*-values are derived from non-parametric statistical methods, i.e. Wilcoxon rank-sum test for within- and Mann-Whitney-U-test for between-subjects comparisons.

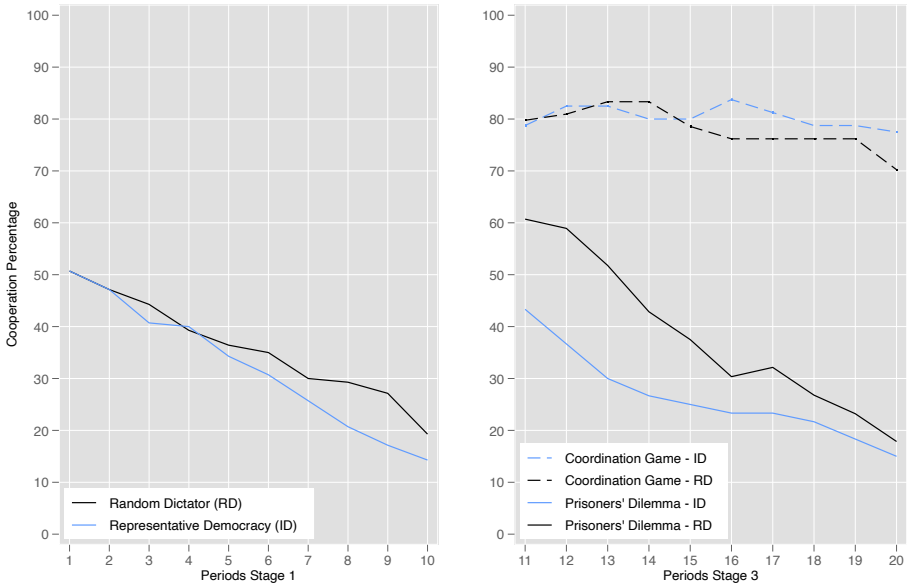
Average cooperation in the first stage, in which all groups played the regular prisoners' dilemma, amounts to 32 percent in the ID treatment and 36 percent in RD

[5] 5 subjects chose the option "other or prefer not to say". As no further information about the subjects' gender who chose the "other" option is available, they are excluded from subsequent analyses of gender effects.

[6] On average female subjects gave correct answers to 1.5 questions while males answered 2.0 questions correctly (p-value < 0.001). This pattern of gender differences is curious but consistent with previous studies using the cognitive reflection test (Frederick, 2005, p.37).

(see left panel of Figure 2.3). The difference is rather small and not statistically significant. The general pattern of positive but decreasing cooperation is well-known from previous experiments on the prisoners' dilemma (Cooper et al., 1996).

FIGURE 2.3: Cooperation Rates – Stages 1 and 3



In the vote stage, 63 percent of subjects in ID and 65 percent in RD chose to modify the payoffs in case they became group leader ($p = 0.709$). Table 2.3 shows Probit and linear probability models (LPM) of voting for modification regressed on variables covering experiences from stage 1 and personal characteristics. Cooperation is positively related to a preference for the coordination game and significant in every specification. A player who cooperated in all of the first ten rounds has a probability to favor the coordination game that is almost 50 percentage points higher than that of someone who did not cooperate at all. Furthermore, partners' cooperation has a significantly negative influence, which is intuitive since a subject with cooperative group members in stage 1 would see less necessity to switch to the coordination game. Cognitive reflection, as measured by the variable *logic*, is significantly positively correlated with a preference for modification. A subject who was able to answer all three logic questions has a probability of voting for modification that is 30 percentage points higher compared to one who gave no correct answer. The treatment dummy variable is not significant, which supports the theoretical conjecture that modification preferences are formed independent of treatment. Overall, the data lends support to Hypothesis 1 implying that cooperative actors self-select into matching institutions.

Result 4 *Cooperative individuals and those with a higher cognitive ability have an increased preference for the coordination game.*

TABLE 2.3: Individual Determinants of Institutional Preferences

Dependent variable: preference for the coordination game

	LPM		Probit	
	(1)	(2)	(3)	(4)
Own cooperation	0.048***	0.047***	0.135***	0.148***
	(0.013)	(0.012)	(0.040)	(0.041)
Partners' cooperation	-0.037***	-0.036***	-0.103**	-0.108***
	(0.013)	(0.019)	(0.040)	(0.041)
Logic		0.102***		0.300***
		(0.027)		(0.081)
Random Dictator		0.010		0.040
		(0.051)		(0.151)
Female		-0.062		-0.226
		(0.058)		(0.185)
Age		-0.013**		-0.040**
		(0.005)		(0.017)
Economics		0.064		0.175
		(0.052)		(0.161)
Constant	0.602***	0.762***	0.264**	0.836*
	(0.046)	(0.161)	(0.124)	(0.502)
N	280	276	280	276
R^2	0.054	0.155		
Pseudo R^2			0.042	0.128

Note: Dependent variable equal to one if a subject chose the coordination game. Own and partners' cooperation range from 0 to 10 for each round of cooperation in stage 1. Logic ranges from 0 to 3 for each correct cognitive reflection question. Standard errors (in parentheses) clustered at group level.
* $p < 0.10$ ** $p < 0.05$, *** $p < 0.01$.

In the representative democracy treatment, nearly half of the participants (43 percent) declared in the questionnaire they had voted for a group member because it had appeared cooperative in the first stage. But in fact the elected representatives cooperated *less* than the other players in stage 1 (on average 2.5 versus 3.5 out of 10 rounds, $p = 0.02$). There is neither a significant difference in modification preference nor in cooperation in stage 3 between the elected representatives and the rest of the subjects ($p = 0.31$ and $p = 0.79$, respectively).

Result 5 *Elected representatives are not more cooperative than other subjects.*

Regarding the difference in leadership behavior between the treatments, contrary to Hypothesis 3, the randomly appointed leaders in RD cooperate on average 6.3 out of 10 times in Stage 3. The elected representatives in ID cooperate 5.8 out of 10 times and the difference is not significant ($p = 0.36$). The results thus fail to confirm the findings by (Drazen and Ozbay, 2019). However, leaders in both treatments considerably increase their average cooperation rate between the two stages. Cooperation of leaders in Stage 3 goes up by 90 percent in the RD treatment and by 130 percent in ID.

Result 6 *Elected representatives are not more cooperative than randomly appointed group leaders.*

Table 2.4 shows the outcome of the vote stage after the computer intervention. The leaders' decision was considered for 31 groups in total. Out of those, 12 leaders in ID and 10 in RD modified the payoffs to play the coordination game. The remaining 9 chose to stick to the prisoners' dilemma. In both treatments cooperation rates are higher (see right panel of Figure 2.3 and more stable under the modified payoffs than in the prisoners' dilemma ($p < 0.01$, two-sided t-test). The institution itself is therefore effective in fostering cooperative behavior as subjects respond to the incentives of the different games and mostly coordinate on the more efficient equilibrium. However, the pure difference between the two games is potentially biased by self-selection and not informative about the effect of democratic policy selection.

To control for a selection bias the analysis of cooperation is conditioned on the four possible vote stage outcomes as well as a subject's institutional preference. Table 2.4 gives the cooperation rates directly after the vote stage in addition to the number of observations. There is no statistical difference between the share of subjects preferring the institution in EndoMod and ExoMod ($p = 0.8$ in both treatments). But the share of yes-voters is significantly lower in the two Not conditions of both treatments ($p < 0.01$ in RD and $p = 0.04$ in ID).

Behavior in Round 11 is of particular interest because it captures the first round after the vote stage and thus the first round in which the institutional change has come into effect. In ID, average cooperation differs by around 17 percentage points between EndoMod and ExoMod: cooperation is higher if the payoff modification policy is democratically introduced by an elected representative (see Tables 2.5 and 2.6). While all vote stage outcomes are significantly different from zero, there are not always significant differences between the outcomes, especially when controlling for modification preferences. The difference between EndoMod and ExoMod is significant at the 10-percent level.

ExoNot induces cooperation rates that are more than twice as high as in EndoNot in the ID treatment ($p = 0.009$). This phenomenon demonstrates the flip side of democracy: endogenously chosen non-cooperative institutions lead to large-scale defection. Behavior under ExoNot is statistically indistinguishable from EndoMod

TABLE 2.4: Cooperation Rates By Vote Outcome

Cooperation rate in round 11 (in percent)				
	EndoMod	*EndoNot*	*ExoMod*	*ExoNot*
Individual vote				
	Representative Democracy (ID)			
No	64.3	23.1	40.0	46.7
	[14]	[13]	[10]	[15]
Yes	94.1	27.3	81.8	61.9
	[34]	[11]	[22]	[21]
All	85.4	25.0	68.8	55.6
	[48]	[24]	[32]	[36]
	Random Dictator (RD)			
No	80.0	22.2	66.7	44.4
	[10]	[9]	[12]	[18]
Yes	80.0	66.7	84.4	84.6
	[30]	[3]	[32]	[26]
All	80.0	33.3	79.5	68.2
	[40]	[12]	[44]	[44]

Note: Number of subjects in brackets.

in ID, even though there are two different payoff schemes at play. Thus, the implementation procedure through the elected representative versus the computer is of great relevance to subjects.

In RD no significant effect can be found between EndoMod and ExoMod ($p = 0.96$); the behavior is virtually identical here. The only vote stage outcome that stands out in RD is EndoNot. Again, we can observe a negative democracy premium in the prisoners' dilemma between EndoNot and ExoNot ($p = 0.03$). When considering the entire third stage, treatment differences become even more pronounced: the difference between EndoMod and ExoMod is highly significant in ID ($p < 0.001$), but not in RD ($p = 0.82$).

Next, I estimate the size of the democracy premium: the increase in cooperation that cannot be explained by policy change or group composition. The democracy premium is the residual in cooperation differences after controlling for the instrumental effect of the payoff modification itself, potential selection effects, and information effects from the voting procedure.

The identification strategy by DFP (2010) breaks down the total policy effect into a selection effect, the exogenous treatment effect, and the democracy premium. This is done by using weighted averages of the individual cooperation rates and voter shares in round 11 that are provided in Table 2.4 (the complete calculation can be found in the appendix). The weighting by modification preferences accounts for differences in group composition: as modification preference and cooperation are

TABLE 2.5: The Effect of Democracy

Dependent Variable: Cooperation in Round 11				
	(1) ID	(2) RD	(3) ID	(4) RD
EndoMod	0.854***	0.800***		
	(0.063)	(0.069)		
EndoNot	0.250***	0.333***		
	(0.090)	(0.126)		
ExoMod	0.688***	0.795***		
	(0.078)	(0.066)		
ExoNot	0.556***	0.682***		
	(0.073)	(0.066)		
EndoModn			0.643***	0.800***
			(0.114)	(0.133)
EndoNotn			0.231*	0.222
			(0.118)	(0.140)
ExoModn			0.400***	0.667***
			(0.135)	(0.121)
ExoNotn			0.467***	0.444***
			(0.110)	(0.099)
EndoMody			0.941***	0.800***
			(0.073)	(0.077)
EndoNoty			0.273**	0.667***
			(0.129)	(0.243)
ExoMody			0.818***	0.844***
			(0.091)	(0.074)
ExoNoty			0.619***	0.846***
			(0.093)	(0.082)
N	140	140	140	140
R^2	0.71	0.74	0.73	0.77

Note: OLS results, dependent variable equal to one for cooperation. Independent variables are binary indicators for vote stage outcomes. Models estimated without a constant. Suffixes denote interactions with individual preferences for (-y) or against (-n) modification. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.6: The Effect of Democracy: p-values

	(1) ID	(2) RD	(3) ID	(4) RD
EndoNot = ExoNot	0.009***	0.015**		
EndoMod = ExoMod	0.099*	0.962		
EndoMod = EndoNot	0.000***	0.001***		
ExoMod = ExoNot	0.219	0.223		
EndoNotn = ExoNotn			0.146	0.198
EndoModn = ExoModn			0.171	0.460
EndoModn = EndoNotn			0.013**	0.003***
ExoModn = ExoNotn			0.702	0.485
EndoNoty = ExoNoty			0.031**	0.683
EndoMody = ExoMody			0.293	0.683
EndoMody = EndoNoty			0.00***	0.601
ExoMody = ExoNoty			0.128	0.983

Note: p-values of Wald tests for differences between vote stage outcomes based on regression results reported in Table 2.14.

positively correlated, a group with more yes-voters can be expected to have higher cooperation rates. Therefore, the estimation approach by DFP (2010) holds voter shares constant across vote stage outcomes. Table 2.7 gives on overview of the treatment effects in ID and RD.

The total effect of the policy – the change from one game to another – is given by the difference between EndoMod and EndoNot and amounts to 60 percentage points in ID and 47 in RD. The effect of endogenously switching to the coordination game is thus substantially larger for the representative democracy. The selection effect captures the higher cooperation that would be observed in the EndoNot condition if the share of yes-voters was the same as in the EndoMod groups. The difference in the proportion of player types that leads to differences in behavior beyond the differing treatment is small in ID and much larger in RD. The change in cooperation caused by an exogenous payoff modification is given by the exogenous treatment effect which measures the difference between the two exogenous conditions. By keeping the proportion of yes- and no-voters as in the endogenous estimate and using the cooperation rates from ExoNot and ExoMod, it is estimated at 12 percentage points in ID and 5 in RD. Lastly, the hitherto unaccounted part of the total policy effect gives the democracy premium which accounts for 47 percentage points in ID and 19 in RD. The representative democracy induces an inexplicable increase in cooperation that is more than double than that of the random dictator. This democracy premium in ID is to a large extent driven by a pronounced reaction from the no-voters, who cooperate much more after endogenous modification than with the exogenous modification. The recommendation towards cooperation from the elected group representative has a much stronger influence on behavior than modification through the computer. The more legitimate decision-making process of democratic choice seems to be especially powerful in combination with the focal figure of an elected representative. The RD democracy premium is positive as well, but it should be kept in mind that the difference between EndoMod and ExoMod is

not significant. The result is rather driven from the opposite direction: a significant *decrease* in cooperation in EndoNot. So if anything, as opposed to the elected representative, the endogenous decision by a random dictator can have a negative impact on cooperation. Subjects do not seem to appreciate an unelected – and therefore potentially illegitimate – leader.

TABLE 2.7: Treatment Effects – Decomposing Cooperation Rates

	Representative Democracy (ID)	Random Dictator (RD)
Total policy effect	60	47
Selection effect	1	22
Exogenous treatment effect	12	5
Democracy premium	47	19

As a robustness check, an additional estimation approach is conducted: Dal Bó, Foster, and Kamei (2019) introduce an alternative identification strategy to estimate the size and significance of the effect of endogenous decision-making by using weighted averages of voting behavior instead of the individual voting decisions. The cooperation rates in the endogenous conditions are reweighted according to the yes- and no-voter shares of the exogenous conditions. In both treatments the subjects in favor of the modification are overrepresented in EndoMod compared to the exogenous conditions (70.8% yes-voters against 63.2% in ID and 75.0% against 65.9% in RD). The voter shares from the exogenous conditions are used as weights for the endogenous conditions to account for the selection effects arising from the democratic choice. The difference between the weighted average cooperation rates gives the democracy effect (see Table 2.8). In ID, it is positive for the modified payoffs – albeit smaller than in the decomposition analysis discussed above – and significant at the 10-percent level. The democracy effect in the modified payoffs in RD is economically and statistically insignificant. There is, however, a highly significant *negative* democracy effect in both treatments: subjects are much less cooperative if the payoffs are endogenously unmodified, and the effect is almost twice as large in ID than in RD. The alternative identification strategy thus yields qualitatively the same result as the decomposition: There is a strong effect of endogenous procedures if the group leader was democratically elected.

The differences between the vote stage outcomes become even more pronounced over the course of the third stage, where behavior is self-reinforcing within groups. Figures 2.4 and 2.5 show individual cooperation rates for all four vote stage results separated by individual voting behavior and treatment over the course of the entire experiment. It can be observed from the first panel of Figure 2.4 that in ID the endogenous institution has a considerable effect on those who preferred modification and leads to almost full cooperation. For both conditions the change of payoff structure results in a striking increase in willingness to cooperate. The rates are

TABLE 2.8: Treatment Effects – Weights-Based Analysis

Weighted average cooperation rates in Round 11					
Treatment	Payoffs	Endo	Exo	Democracy Effect	Standard Error (p-Value)
ID	Mod	83.2 [48]	68.8 [32]	14.4	8.28 (0.082)
	Not	25.7 [24]	55.6 [36]	-29.8	8.44 (<0.001)
RD	Mod	80.0 [40]	79.5 [44]	0.5	6.14 (0.941)
	Not	51.5 [12]	68.2 [44]	-16.7	7.06 (0.018)

Note: Number of subjects in brackets. Bootstrapped standard errors.

much lower without modification, especially if this was endogenously determined. The lower panel shows only individuals who were against the payoff modification in ID. The no-voters who received the coordination game through their representative drastically change their behavior and display quite stable cooperation rates. In the exogenous condition both games induce remarkably similar behavior suggesting that the democratic procedure has a stronger behavioral impact than the change in monetary incentives. Such a strong effect on no-voters is in line with the findings of Gallier (2020), but in contrast with DFP (2010), where the democracy premium was driven by those in favor of the policy. Figure 2.5 corroborates the result that the random dictator's implementation of the coordination game did not increase cooperation more than the exogenous modification: the EndoMod cooperation rates are for both voter types for a large part closely below those in ExoMod. The mere authority to implement a policy decision does not seem to activate compliance in the absence of democratic legitimacy.

Result 7 *Cooperation is higher if the policy is introduced by a representative (ID) and the size of the democracy premium is substantial. Payoff modification by an unelected group leader (RD) does not increase cooperation.*

The treatment difference between ID and RD is informative regarding the source of the democracy premium as conjectured in Hypotheses 5a and b. The evidence from the experiment is clear in this regard: the positive increase in cooperation as a response to the group leader being considered is much larger in ID. In the RD treatment, the difference between EndoMod and ExoMod is small and statistically insignificant.

FIGURE 2.4: Cooperation Rates – Representative Democracy (ID)

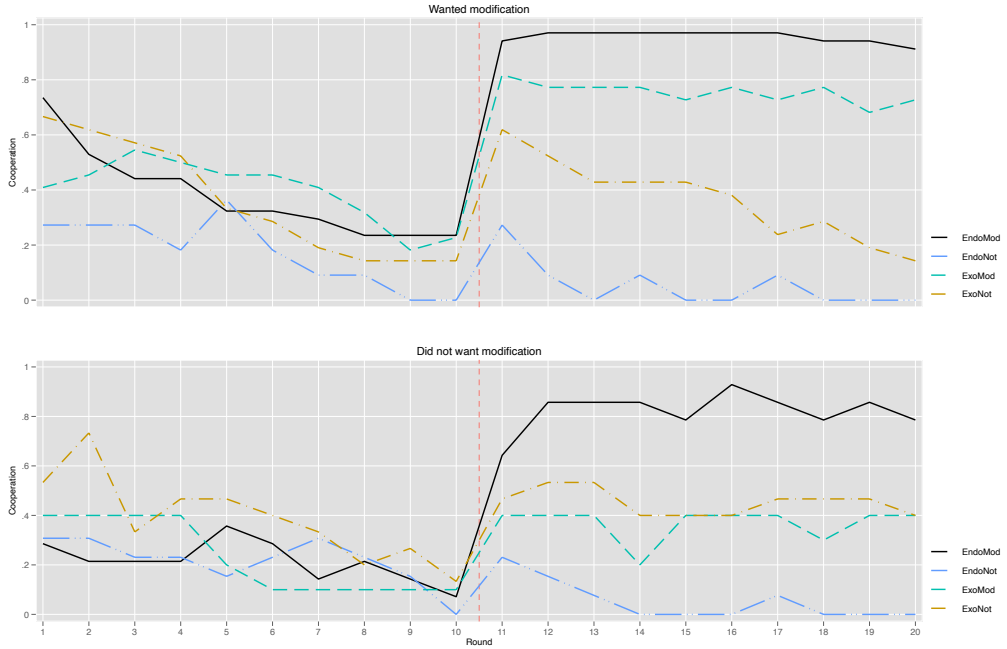
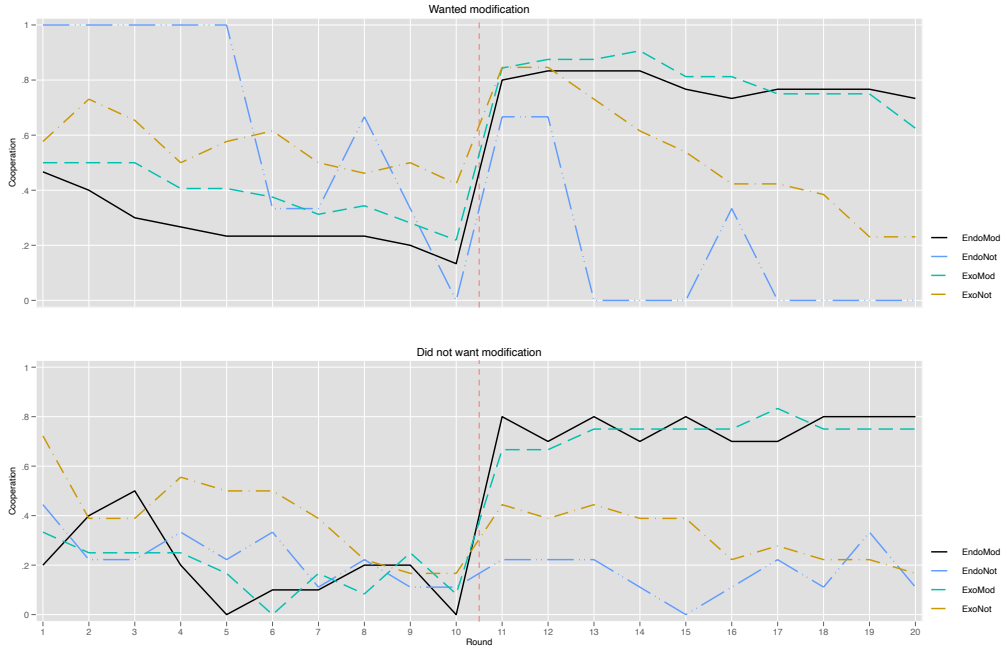


FIGURE 2.5: Cooperation Rates – Random Dictator (RD)



The legitimacy created by the election process seems to be a necessary condition for the democracy premium; authority alone cannot foster cooperation in the same way. Therefore, we reject Hypothesis 5.b and support 5.a instead: perceived legitimacy is the transmission channel of the democracy effect.

Result 8 *Representative democracy goes along with a significantly larger democracy premium than a randomly appointed leader.*

2.4.3 Welfare Implications

A natural next question is whether the the democracy premium corresponds to an increase in societal welfare. Since the exact utility functions of the subjects are unknown, we will restrict attention to monetary payoffs at first. The welfare consequences of the institutional change from one game to the other are not obvious exante. The payoff modification to the coordination game, on the one hand, makes mutual cooperation more attainable and can thus increase overall payoffs. Recalling the payoffs from Table 2.1, mutual cooperation is the efficient outcome and gives 50 points to each player. On the other hand, if coordination fails, the penalty on unilateral defection decreases the earnings in the coordination game to 48 points as opposed to the prisoners' dilemma's deviation payoff of 60 points.

Comparing average earnings in Stage 3, we find that in both treatments average payoffs are lower in the prisoners' dilemma (Table 2.9). The highest average is realized in the EndoMod condition in ID and in ExoMod in RD. However, the maximum payoff earned by one subject over the entire experiment with 53 points on average took place in the ExoMod outcome of ID. All in all, welfare is positively affected by the payoff change in both treatments. But the endogenous modification is the most efficient condition only in ID. Since average payoffs are always higher in the coordination game, it would be a natural conclusion for a social planner to circumvent the voting procedure altogether, which, after all, bears the risk of players choosing the payoff-dominated prisoners' dilemma. Instead, one could simply assign the coordination game to every group. But this comes at a cost: the highest possible cooperation rates – and earnings – are only realized after the endogenous choice. Furthermore, the monetary analysis omits some important aspects of legitimate procedures, which are at the heart of the democracy premium.

For example, drawing on arguments put forward by Thibaut and Walker (1975) and Sen (1995), Frey, Benz, and Stutzer (2004) introduce procedural utility to incorporate preferences about the processes that lead to instrumental outcomes into individual utility functions. The ability to exercise political participation is one source of procedural utility (Frey and Stutzer, 2005). Thus, it is plausible that subjects in the experiment presented in this paper derive higher utility from having their elected representatives considered, which is something that a purely monetary welfare analysis cannot adequately capture.

TABLE 2.9: Average Earnings per Vote Stage Outcome

Treatment	**ID**		**RD**	
Game	Prisoners' Dilemma	Coordination Game	Prisoners' Dilemma	Coordination Game
Implementation				
Endogenous	40.4	48.5	41.7	46.7
Exogenous	44.0	45.3	44.4	47.0
Total	42.6	47.2	43.8	46.9

Note: Average earnings calculated over all ten rounds in stage 3.

2.5 Discussion and Conclusion

The paper uses an economic experiment to quantify the influence that decision-making processes have on cooperative behavior. The effect of a decision made by an elected representative is contrasted with that of a randomly chosen group leader. Subjects are presented with the possibility of changing their payoff structure from a prisoners' dilemma into a coordination game that makes cooperation incentive-compatible. The effect of the procedure that leads to the payoffs being modified or not is given by the extent of cooperative behavior that follows the decision. The experiment uses the identification strategy developed by DFP (2010). A randomization mechanism allows for the comparison of subjects with the same preferences, information, and incentive structure who only differ in how the incentive structure was implemented: by the group leader or by the computer. Additionally, the treatments compare the effects of elected and randomly chosen leaders.

To summarize results, the majority of subjects in both treatments prefer to modify the payoffs, and cooperative players favor the modification more, suggesting that it is appropriate to control for self-selection. In the representative democracy treatment, subjects prefer to elect pro-social representatives. Still, factually these do not behave significantly differently compared to the rest of the subjects or the randomly appointed leaders in the random dictator treatment. The findings stress the importance of procedural legitimacy over elections as selection devices: not who is the leader matters, but how the leader came into office influences behavior. Moreover, subjects cooperate more if the payoff modification is democratically introduced. In contrast to DFP (2010), the impact of the democratic policy selection is especially large for those subjects who initially did not want to introduce the modification. There is no democracy premium in the random dictator treatment. The results show that subjective legitimacy is a driving force of the democracy premium, which an unelected group leader cannot deliver.

The relevance of the results is twofold. First, the behavioral effects of institution formation are relevant for evaluating any experimental treatment effects in which subjects are assigned to different institutions. Second, the results carry policy implications. A policy that works well in one place cannot automatically be assumed to achieve similar results in another context. Further, representation is multi-faceted

and how a leader is chosen determines the success of their policies to a great extent. The results from the ID treatment are reassuring in this sense: representative democracy is a widespread form of government and seems to have the largest positive impact on cooperative behavior compared to direct democracy and sortition. The sortition mechanism modeled in the RD treatment has regained popularity in recent years, mostly from grassroots movements and citizens' initiatives, especially for environmental policy questions (Dryzek and Tucker, 2008; Lorent, 2019; Zimmer, 2021). Even though it has theoretical advantages, i.e., little proneness to corruption, the experimental results presented here imply that one should be cautious regarding the procedural legitimacy and thus the effectiveness of these mechanisms.

Can the democracy premium be explained by economic theory? Markussen, Putterman, and Tyran (2014) claim that it is rationalizable with the model of inequality aversion by Fehr and Schmidt (1999). Voting is a credible signal of an intention to cooperate that prompts inequality-averse subjects to cooperate in the coordination game (Markussen, Putterman, and Tyran, 2014, p.307). However, the argument has no bite in the experiment presented in this paper because subjects are informed about their representative's intention to modify payoffs even when it is not considered. The signaling component does therefore neither differ between the endogenous and exogenous conditions nor between the two treatments. There is no reason why an elected representative should deliver stronger cues towards the cooperative equilibrium than the randomly appointed leader if we restrict attention to inequality aversion. Even when the players are assumed to be not purely maximizing their own payoffs, the procedure itself is not sufficiently consequential to account for the democracy premium. If preferences about political participation enter the utility functions, e.g., in the form of procedural utility Frey, Benz, and Stutzer (2004) this would create a level effect on subjects' utility in the endogenous conditions. But to create the democracy premium, the procedural preferences have to interact with the treatment conditions in a way that creates differences in *behavior*, not utility. A "warm glow" feeling of political participation has to induce players to cooperate if they were considered but make them defect if they were not. Dannenberg and Gallier (2020) suggest that the endogenous and collective implementation of an institution may evoke feelings of group identity, which can be a powerful activator of social-preferences (Akerlof and Kranton, 2000; Chen and Li, 2009). In a group-contingent social preference model, group identity can influence equilibrium selection in coordination games (Chen and Chen, 2011). Such a model could serve to explain the democracy premium under the assumption that the interaction between subjects in the first stages is not sufficient to induce group identity in the exogenous conditions, and moreover, that endogenously refusing the payoff change has adverse effects on group identity. The perhaps most promising avenue for further theoretical research on the democracy premium is Psychological Game Theory (PGT) (Geanakoplos, Pearce, and Stacchetti, 1989; Battigalli and Dufwenberg, 2009).

PGT formally incorporates belief-dependent motivations into game theory. One especially noteworthy application is guilt aversion (Battigalli and Dufwenberg, 2007). In a guilt aversion framework, disutility is created from a failure to live up to others' expectations. The model fits well with the democracy premium if the choice to modify the payoffs is a statement of intent to cooperate. Under endogenous modification, a deviation from mutual cooperation would then be seen as "letting the other player down". It is up to further research should develop a theoretical synthesis of the numerous experimental studies on the democracy premium and their ambiguous findings. One important inconsistency in the literature on the democracy premium is who is affected most by the the democratic process: the ones in favor or the ones opposing the institutional change. Theoretical explanations hinge on this as much as policy implications derived from the experimental insights.

Limitations of the abstract experimental design presented here are that many essential features of representative democracies are excluded in the ID treatment. There is no running for elections, neither pandering nor accountability, and no rent for the elected politician. As such, the external validity of the results can be further improved. Incorporating the randomization mechanism into more complex experiments promises to deliver unbiased estimates of various kinds of endogenous treatment effects. Possible extensions to the study are giving more power to the representative, e.g., by letting her decide on the strategies of the citizens. Repeated elections that create accountability, campaigning of candidates, and preference heterogeneity are further relevant factors in representative democracies to potentially include.

Bibliography

Akerlof, George A and Rachel E Kranton (2000). "Economics and Identity". In: *The Quarterly Journal of Economics* 115.3, pp. 715–753.

Alizada, Nazifa, Rowan Cole, Lisa Gastaldi, Sandra Grahn, Sebastian Hellmeier, Palina Kolvani, Jean Lachapelle, Anna Lührmann, Seraphine F. Maerz, Shreeya Pillai, and Staffan I. Lindberg (2021). *Autocratization Turns Viral. Democracy Report 2021*. Tech. rep.

Andreoni, James and Laura K Gee (2012). "Gun for hire: Delegated enforcement and peer punishment in public goods provision". In: *Journal of Public Economics* 96.11-12, pp. 1036–1046.

Bardhan, Pranab (2000). "Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India". In: *Economic Development and Cultural Change* 48.4, pp. 847–865.

Barnes, Jonathan (1984). *Complete works of Aristotle, Volume 1: The revised Oxford translation*. Vol. 1. Princeton University Press.

Battigalli, Pierpaolo and Martin Dufwenberg (2007). "Guilt in games". In: *American Economic Review* 97.2, pp. 170–176.

— (2009). "Dynamic psychological games". In: *Journal of Economic Theory* 144.1, pp. 1–35.

Becker, Gary S. (1976). *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Black, Sandra E. and Lisa M. Lynch (2001). "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity". In: *Review of Economics and Statistics* 83.3, pp. 434–445.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg Registration and Organization Online Tool". In: *European Economic Review* 71.1, pp. 117–120.

Bonin, John P., Derek C. Jones, and Louis Putterman (1993). "Theoretical and Empirical Studies of Producer Cooperatives: Will Ever the Twain Meet?" In: *Journal of Economic Literature* 31.3, pp. 1290–1320.

Brandts, Jordi, David J Cooper, and Roberto A Weber (2015). "Legitimacy, communication, and leadership in the turnaround game". In: *Management Science* 61.11, pp. 2627–2645.

Brandts, Jordi, Werner Güth, and Andreas Stiehler (2006). "I want YOU! An experiment studying motivational effects when assigning distributive power". In: *Labour Economics* 13.1, pp. 1–17.

Castillo, Jose G, Zhicheng Xu, Ping Zhang, and Xiacheng Xu (2017). "Institutional Legitimacy and Public Goods Games: A Laboratory Experiment on the Distribution of Sanctioning Power". In: *mimeo*.

Cavalcanti, Carina, Felix Schläpfer, and Bernhard Schmid (2010). "Public Participation and Willingness to Cooperate in Common-pool Resource Management: A Field Experiment with Fishing Communities in Brazil". In: *Ecological Economics* 69.3, pp. 613–622.

Chen, Roy and Yan Chen (2011). "The potential of social identity for equilibrium selection". In: *American Economic Review* 101.6, pp. 2562–89.

Chen, Yan and Sherry Xin Li (2009). "Group identity and social preferences". In: *American Economic Review* 99.1, pp. 431–57.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1996). "Cooperation Without Reputation: Experimental Evidence from Prisoner's Dilemma Games". In: *Games and Economic Behavior* 12.2, pp. 187–218.

Corazzini, Luca, Sebastian Kube, Michel André Maréchal, and Antonio Nicolo (2014). "Elections and Deceptions: An Experimental Study on the Behavioral Effects of Democracy". In: *American Journal of Political Science* 58.3, pp. 579–592.

Dal Bó, Pedro (2014). "Experimental evidence on the workings of democratic institutions". In: *Institutions, Property Rights, and Economic Growth: The Legacy of Douglass North*. Ed. by Sebastián Galiani and Itai Sened. Cambridge: Cambridge University Press, pp. 266–288.

Dal Bó, Pedro, Andrew Foster, and Kenju Kamei (2019). *The Democracy Effect: a weights-based identification strategy*. Tech. rep. National Bureau of Economic Research.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy". In: *American Economic Review* 100.5, pp. 2205–29.

Dannenberg, Astrid and Carlo Gallier (2020). "The choice of institutions to solve cooperation problems: a survey of experimental research". In: *Experimental Economics* 23, 716—749.

Dannenberg, Astrid, Corina Haita-Falah, and Sonja Zitzelsberger (2020). "Voting on the threat of exclusion in a public goods experiment". In: *Experimental Economics* 23, pp. 84–109.

Drazen, Allan and Erkut Y Ozbay (2019). "Does "being chosen to lead" induce non-selfish behavior? Experimental evidence on reciprocity". In: *Journal of Public Economics* 174, pp. 13–21.

Dryzek, John S and Aviezer Tucker (2008). "Deliberative innovation to different effect: Consensus conferences in Denmark, France, and the United States". In: *Public Administration Review* 68.5, pp. 864–876.

Fearon, James D, Macartan Humphreys, and Jeremy M Weinstein (2011). "Democratic Institutions and Collective Action Capacity: Results from a Field Experiment in Post-conflict Liberia". In: *APSA 2011 Annual Meeting Paper*.

Fehr, Ernst and Klaus M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation". In: *Quarterly Journal of Economics* 114.3, pp. 817–868.

Frederick, Shane (2005). "Cognitive reflection and decision making". In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Frey, Bruno S., Matthias Benz, and Alois Stutzer (2004). "Introducing Procedural Utility: Not Only What, But Also How Matters". In: *Journal of Institutional and Theoretical Economics* 160.3, pp. 377–401.

Frey, Bruno S and Alois Stutzer (2005). "Beyond outcomes: measuring procedural utility". In: *Oxford Economic Papers* 57.1, pp. 90–111.

Gallier, Carlo (2020). "Democracy and compliance in public goods games". In: *European Economic Review* 121, p. 103346.

Geanakoplos, John, David Pearce, and Ennio Stacchetti (1989). "Psychological games and sequential rationality". In: *Games and Economic Behavior* 1.1, pp. 60–79.

Grossman, Guy and Delia Baldassarri (2012). "The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda". In: *American Journal of Political Science* 56.4, pp. 964–985.

Ham, Carolien van, Jacques JA Thomassen, Kees Aarts, and Rudy B Andeweg (2017). *Myth and reality of the legitimacy crisis: Explaining trends and cross-national differences in established democracies*. Oxford University Press.

Hamman, John R., Roberto A. Weber, and Jonathan Woon (2011). "An Experimental Investigation of Electoral Delegation and the Provision of Public Goods". In: *American Journal of Political Science* 55.4, pp. 738–752.

Kamei, Kenju (2016). "Democracy and resilient pro-social behavioral change: an experimental study". In: *Social Choice and Welfare* 47.2, pp. 359–378.

— (2017). "Collective Selection of Representatives and the Democracy Effect". In: *mimeo*.

Kamei, Kenju, Louis Putterman, and Jean-Robert Tyran (2015). "State or Nature? Endogenous Formal Versus Informal Sanctions in the Voluntary Provision of Public Goods". In: *Experimental Economics* 18.1, pp. 38 –65.

Locke, John (1983). *The Second Treatise of Government*. Stuttgart: Reclam.

Lorent, Pascal (2019). "La Communauté germanophone se dote d'une assemblée citoyenne". In: *Le Soir*. Accessed: 2021-04-11.

Marcin, Isabel, Pedro Robalo, and Franziska Tausch (2019). "Institutional endogeneity and third-party punishment in social dilemmas". In: *Journal of Economic Behavior & Organization* 161, pp. 243–264.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2014). "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes". In: *Review of Economic Studies* 81.1, pp. 301–324.

Mechtenberg, Lydia and Jean-Robert Tyran (2019). "Voter Motivation and the Quality of Democratic Choice". In: *Games and Economic Behavior* 116.

Olken, Benjamin A. (2010). "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia". In: *American Political Science Review* 104.02, pp. 243–267.

Ostrom, Elinor (1991). *Governing the Commons*. Cambridge: Cambridge University Press.

Ostrom, Elinor, James Walker, and Roy Gardner (1992). "Covenants with and without a sword: Self-governance is possible". In: *American Political Science Review* 86.2, pp. 404–417.

Rivas, M. Fernanda and Matthias Sutter (2011). "The Benefits of Voluntary Leadership in Experimental Public Goods Games". In: *Economics Letters* 112.2, pp. 176–178.

Schmidt, Vivien A. (2013). "Democracy and Legitimacy in the European Union Revisited: Input, Output and 'Throughput'". In: *Political Studies* 61.1, pp. 2–22.

Sen, Amartya (1995). "Rationality and social choice". In: *American Economic Review* 85.1, pp. 1–24.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher (2010). "Choosing the carrot or the stick? Endogenous Institutional Choice In Social Dilemma Situations". In: *The Review of Economic Studies* 77.4, pp. 1540–1566.

Thibaut, John W and Laurens Walker (1975). *Procedural Justice: A Psychological Analysis*. Hillsdale: Lawrence Erlbaum Associates.

Towfigh, Emanuel V, Sebastian J Goerg, Andreas Glöckner, Philip Leifeld, Aniol Llorente-Saguer, Sophie Bade, and Carlos Kurschilgen (2016). "Do direct-democratic procedures lead to higher acceptance than political representation?" In: *Public Choice* 167.1-2, pp. 47–65.

Tyran, Jean-Robert and Lars P. Feld (2006). "Achieving Compliance When Legal Sanctions Are Non-deterrent". In: *The Scandinavian Journal of Economics* 108.1, pp. 135–156.

Vollan, Björn, Andreas Landmann, Yexin Zhou, Biliang Hu, and Carsten Herrmann-Pillath (2017). "Cooperation and authoritarian values: An experimental study in China". In: *European Economic Review* 93, pp. 90–105.

Zimmer, Stefan (2021). *Klimaneustart Berlin*. `https://klimaneustart.berlin/klima-buergerinnenrat/`. Accessed: 2021-04-11.

B Appendix to Chapter 2

B.1 Treatment Effect Calculation

Following DFP (2010) the policy effects in round 11 can be calculated as follows.

Representative Democracy

Total policy effect = Selection effect + Exogenous treatment effect + Democracy premium

Total policy effect: $[64.3(14/48) + 94.1(34/48)] - [23.1(13/24) + 27.3(11/24)] = 60.38$.

Selection effect: 23.1(14/48 - 13/24) + 27.3(34/48 - 11/24)] = 1.05.

Exogenous treatment effect: (14/48)(40.0 - 46.7) + (34/48)(81.8 - 61.9) = 12.14.

Democracy premium: 59.33 - 12.14 = 47.19.

Random dictator

Total policy effect [80.0(10/40) + 80.0(30/40)] - [22.2(9/12) + 66.7(3/12)] = 46.68.

Selection effect: 22.2(10/40 - 9/12) + 66.7(30/40 - 3/12)] = 22.25.

Exogenous treatment effect: (10/40)(66.7 - 44.4) + (30/40)(84.4 - 84.6) = 5.425.

Democracy premium: 24.43 - 5.43 = 19.

B.2 Robustness Check: Subsample Analysis

This section restricts the attention to subjects whose play in the prisoners' dilemma included more than one action, i.e. excludes those who either always or never cooperate in stage 1. The remaining 89 subjects in ID and 98 in RD can be classified as *conditional cooperators* (Table 2.10).

TABLE 2.10: Number of Subjects per Subsample

Treatment	Defector	Conditional Cooperator	Unconditional Cooperator
ID	39	89	12
RD	32	98	10

Note: Classification according to cooperation in stage 1.

In a regression of cooperative behavior in stage 3 on vote stage outcome, payoff modification preference, and previous cooperation for the restricted samples, the results from the main analysis still hold (Table 2.11): cooperation in EndoMod is significantly higher than in ExoMod, which is the baseline category, but only in the ID treatment. In the case of the randomly appointed leader EndoNot is the only significant vote stage outcome out of the two endogenous cases. The effect is large, significant, and negative. The insight that the positive impact of leadership is only evoked by an elected leader is thus corroborated.

TABLE 2.11: The Effect of Democracy on Conditional Cooperators

Dependent variable: Cooperation in Stage 3		
	(1)	(2)
	ID	RD
EndoMod	2.359**	0.264
	(1.135)	(0.899)
EndoNot	-5.544***	-7.255***
	(1.216)	(0.646)
ExoNot	-3.114**	-4.613***
	(1.447)	(0.798)
Modification	1.226**	-0.511
	(0.523)	(0.530)
Cooperation	0.544***	0.369***
in Stage 1	(0.156)	(0.0995)
Constant	4.134**	7.432***
	(1.527)	(0.894)
N	89	98
R^2	0.667	0.527

Note: Cooperation in Stage 1 and 3 is measured as the sum of cooperative actions chosen by a subject. Sample restricted to observations with cooperation in stage 1 between 1 and 9. All other variables are binary vote stage indicators. Standard errors (in parentheses) clustered at group level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.3 Direct Democracy Replication

In addition to the representative democracy and random dictator treatments a replication of the original DFP(2010) study of a direct democracy was conducted. Notably, the replication varies one crucial aspect of DFP (2010): subjects are informed about the vote outcome in their group even in the exogenous conditions. This ensures that it is the democratic procedure itself, and not informational differences, that drive the results. DFP (2010) themselves consider this design variation as a robustness check, but change one more aspect of the design simultaneously: the exogenous modification probability is increased to 90%. The replication attempt conducted here incorporates the informational treatment, but leaves the intervention probability at 50 %, which is the case in DFP (2010)'s main experiment.

The main findings of DFP (2010) are compared to the results of the replication study in the following section. To summarize, the results of DFP (2010) could not be replicated. The democracy premium is of approximately the same magnitude as in DFP (2010), but it is *negative*. Importantly, the estimate is not significantly different from zero and largely driven by four cooperative subjects, who were by chance the only group in the EndoNot outcome. Acknowledging the small sample size in the replication attempt, the results should be treated with caution. It is not suggested that the findings of DFP (2010) are not valid, but that further replication efforts are needed to assess the results' robustness.

Replication Analysis The number of subjects in the replication sample is 92. In stage 1, average cooperation amounts to 33 percent, which is almost twice as high as in DFP (2010). This first finding already suggests that some underlying characteristics might differ between the populations studied here and in DFP (2010). In the vote stage, 65 percent of subjects voted in favor of the payoff modification, compared to only 53 in DFP (2010). Looking at the determinants of voting for modification in Table 2.12. DFP (2010) find that more cooperative subjects are significantly more likely to vote for the modification. At the same time, cooperative behavior experienced by the opposing players in stage 1 decreased the preference for modification. These effects cannot be replicated in the individual regressions with the data from Hamburg reported in column (2) and (4). Neither effect is significant on its own, however, in the combined regression in column (5), the variables take on comparable signs and effect sizes as in DFP (2010).

The outcome of the vote stage can be seen in Table 13 (compare with Table 4 in DFP, 2010, p.2214). It becomes clear that due to the limited sample size the explanatory power of the estimates in the following will be limited. In particular, no group chose the EndoNot condition by a majority of votes.

Table 13 shows the cooperation rates in the first round after the vote stage. Regardless of the vote stage outcome, it can be seen that yes-voters cooperate more than no-voters. Average cooperation is the highest in EndoMod, which corroborates the results found by DFP (2010). The difference between EndoMod and ExoMod in

TABLE 2.12: Cooperation and Preference for Payoff Modification

	(1) DFP	(2) UHH	(3) DFP	(4) UHH	(5) UHH
	Dependent variable: modification preference				
Own cooperation in stage 1	0.47*** (0.161)	0.232 (0.158)			0.633*** (0.238)
Partners' cooperation in stage 1			-0.419** (0.211)	-0.054 (0.174)	-0.574** (0.246)
Constant	0.448*** (0.042)	0.576*** (0.074)	0.608*** (0.048)	0.670*** (0.075)	0.633*** (0.076)
N	276	92	276	92	92
R^2	0.03	0.02	0.01	0.001	0.078

Note: Linear probability model, dependent variable equal to one for a vote to modify payoffs. Columns (1) and (3) report data from DFP (2010). UHH indicates new data from Hamburg. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.13: Individual Cooperation By Vote Outcome

Cooperation rate in Round 11 (in percent)				
Individual vote	*EndoMod*	*EndoNot*	*ExoMod*	*ExoNot*
	Direct Democracy (DD)			
No	66.7 [15]	50.0 [2]	8.3 [12]	33.3 [3]
Yes	91.9 [37]	100.0 [2]	100.0 [12]	[9]
All	84.6 [52]	75.0 [4]	54.2 [24]	[12]

Note: Number of subjects in brackets.

round 11 amounts to 22 percentage points in DFP (2010) and 31 in the replication. Both difference are highly significant ($p < 0.01$). Among the yes-voters in EndoMod and ExoMod, DFP (2010) find a difference of 24 percentage points, I find a *negative* difference of 8 percentage points, which is not statistically significant ($p = 0.31$) The second highest average cooperation rates are observed in EndoNot, which is contradictory to the hypotheses of the democracy premium. It has to be kept in mind, however, that EndoNot is the condition with the lowest number of observations and should be treated with caution.

Tables 2.14 and 2.15 give more detailed information about the results of the vote stage in the original study and the replication. Most vote stage outcomes are significantly different from zero. The effect sizes for EndoMod are even bigger in the replication compared to DFP (2010). However, Table 2.15 shows that the vote stage conditions are largely statistically undistinguishable in the replication. The difference between EndoMod and ExoMod in Hamburg is driven by the no-voters, which runs contrary to DFP (2010), where the democracy premium stems from the yes-voters. This is worthy of further investigation.

TABLE 2.14: The Effect of Democracy

Dependent Variable: Cooperation in Round 11

	(1) DFP	(2) UHH	(3) DFP	(4) UHH	(5) DFP	(6) UHH
EndoMod	0.722***	0.846***				
	(0.050)	(0.060)				
EndoNot	0.175***	0.750***				
	(0.048)	(0.216)				
ExoMod	0.500***	0.542***				
	(0.053)	(0.0881)				
ExoNot	0.150***	0.583***				
	(0.055)	(0.125)				
EndoModn			0.412***	0.667***	0.362***	0.531***
			(0.101)	(0.090)	(0.102)	(0.105)
EndoNotn			0.145**	0.500**	0.050	0.503**
			(0.056)	(0.246)	(0.067)	(0.230)
ExoModn			0.419***	0.083	0.314***	0.0293
			(0.075)	(0.100)	(0.086)	(0.0966)
ExoNotn			0.038	0.333	-0.016	0.172
			(0.082)	(0.201)	(0.084)	(0.194)
EndoMody			0.818***	0.919***	0.719***	0.769***
			(0.056)	(0.057)	(0.063)	(0.068)
EndoNoty			0.576***	1.00***	0.09	0.937***
			(0.072)	(0.246)	(0.090)	(0.231)
ExoMody			0.576***	1.00***	0.431***	0.844***
			(0.072)	(0.100)	(0.082)	(0.104)
ExoNoty			0.235***	0.667***	0.112***	0.521***
			(0.071)	(0.116)	(0.079)	(0.118)
Own cooperation in stage 1					0.618***	0.461***
					(0.139)	(0.172)
Partners'cooperation in stage 1					-0.034***	-0.061
					(0.179)	(0.190)
N	276	92	276	92	276	92
R^2	0.54	0.76	0.57	0.85	0.60	0.871

Note: OLS results, dependent variable equal to one for cooperation. Independent variables are binary indicators for vote stage outcomes. The models have no constant. Suffixes in columns (3)–(6) denote interactions with individual modification preferences for (-y) or against (-n) modification. Columns (1), (3), and (5) report data from DFP (2010). UHH indicates replication data from Hamburg.
Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.15: p-values

	(1) DFP	(2) UHH	(3) DFP	(4) UHH	(5) DFP	(6) UHH
EndoNot = ExoNot	0.732	0.505				
EndoMod = ExoMod	0.003	0.005				
EndoMod = EndoNot	0.000	0.669				
ExoMod = ExoNot	0.000	0.786				
EndoNotn = ExoNotn			0.281	0.601	0.494	0.273
EndoModn = ExoModn			0.952	0.000	0.694	0.000
EndoModn = EndoNotn			0.022	0.526	0.006	0.910
ExoModn = ExoNotn			0.001	0.269	0.003	0.504
EndoNoty = ExoNoty			0.966	0.224	0.834	0.109
EndoMody = ExoMody			0.009	0.485	0.001	0.486
EndoMody = EndoNoty			0.000	0.749	0.000	0.481
ExoMody = ExoNoty			0.001	0.033	0.001	0.027

Note: p-values of Wald tests for differences between vote stage outcomes, based on regression results reported in Table 2.14. Columns (1), (3), and (5) show data from DFP (2010). UHH indicates the replication data from Hamburg.

B.4 Instructions

Welcome to the experimental lab. Please keep in mind that from now on you are not allowed to communicate with anyone other than the lab personnel. If a question arises please show your hand and we will contact you. You must not use a phone, tablet or similar device throughout the entire session. Please note that any act of non-compliance with these rules may lead to your exclusion from all payments. Every decision you will make during the experiment will be treated anonymously and cannot be linked to your identity. Now, please read these instructions carefully and hand them back to the assistants at the end of the experiment.

The following experiment has two parts. You will receive instructions for the second part after the first is completed. Both parts consist of a game that is played for ten rounds. You will earn points in these games; the amount of points you earn depends on your own and on others' choices. At the end of the session one round from each of the two parts will be randomly selected and paid. Points will be converted at a rate of 10 points = 1 €. First of all you are now randomly divided into groups of four. Simultaneously, every player receives a player ID between 1 and 4. Both the group composition as well as all player IDs remain unchanged throughout the entire experiment.
Example: You are player 2 and form a group with the players 1, 3, and 4.

Part 1

In this part you play ten rounds of a game (**Game 1**) together with one of your other three group members. This other player is randomly chosen in every round and you will be notified at the end of the round who your partner was. **In this game you can decide between the options A and B in each round.** Your partner simultaneously chooses one of the options. While you make your decision, you do not know what your partner chooses. Your income in each round of game 1 is calculated in the following way:

If both you and your partner choose option A you both earn 50 points.
If you choose option A and your partner chooses B, then you earn 30 points and your partner earns 60.
If you choose option B and your partner chooses A, then you earn 60 points and your partner earns 30.
If both you and your partner choose option B you both earn 40 points.

After each round you will see the chosen option of your partner and of the other group members on your computer screen. Table 1 gives an overview of your earnings per round in game 1.

Your Choice	Your Partner's Choice	
	A	B
A	50	30
B	60	40

Game 1

Part 2

Part 2 of the experiment starts with a vote. **Every group elects one of their members as their representative in a secret ballot. This representative can decide which game your group will play for ten more rounds.** The choice is between Game 1 (as known from Part 1) and Game 2. In Game 2 you can again choose between options A and B and your income is calculated in the following way:

If both you and your partner choose option A you both earn 50 points.
If you choose option A and your partner chooses B, then you earn 30 points and your partner earns 48.
If you choose option B and your partner chooses A, then you earn 48 points and your partner earns 30.
If both you and your partner choose option B you both earn 40 points.

Your Choice	Your Partner's Choice	
	A	B
A	50	30
B	48	40

Game 2

At first you must now indicate which game you would choose for your group in case you become representative. This decision is secret until the election of the representative is completed. For this you vote for one other group member. You cannot vote for yourself. In case of a tie one of the players with the highest amount of votes is randomly chosen as representative. The representative's choice of game

becomes binding for the entire group. **However, this choice is only implemented with a probability of 50 percent.** If the representative's game is not implemented the computer randomly selects Game 1 or 2. **Both games are equally likely to be chosen in this case.**

You will be informed about who was elected as representative, which game the representative preferred, if this choice was considered and if not which game your group will play in part 2 and **you play this game for ten rounds.** Again, you are informed about your partner's and other group members' choices after each round.

Subsequently we are going to ask you to fill out a short questionnaire, which has no influence on your income, and determine the two rounds relevant for the payout.

Chapter 3

Is the Democracy Premium WEIRD? A Tale of Two Countries

Author Fanny Schories

Abstract I use an economic lab experiment to quantify the democracy premium – an increase in cooperation associated with democratic decision-making – in Egypt and in Germany. The experimental design models a representative democracy and is able to control for confounding factors such as self-selection and information effects. I find large differences between the two culturally different subject pools: German subjects respond strongly to the democratic policy implementation. Egyptians show no positive behavioral response to the decision-making procedure. Varying degrees of religiosity cannot explain the results.

Keywords: Laboratory Experiment, Representative Democracy, Collective Decision-Making, Social Dilemma, Legitimacy.

JEL Classification: C91, D02, D72, Z18.

3.1 Introduction

Democracy is in crisis and populism rises globally. In the Middle East, high aspirations in terms of public political participation became manifest during the Arab Spring, but tumbled into chaos in many instances. In 2019, almost a decade later, citizens took the streets of Beirut again to protest against a corrupted political elite. While the Lebanese society is deeply fractured along sectarian lines, the citizens stand united by their desire to end nepotism and economic mismanagement (Al Jazeera, 2019). Although democratic governance seems universally desired, it remains a fragile ideal too often, easily deployed by populist agendas and material power struggles. Almost every country in the world today claims to honor democratic governance on paper, and the Pew Research Center survey found that "more than half in each of the nations polled consider representative democracy a very or somewhat good way to govern their country" (Wike et al., 2017, p.3), but the extent to which it is de facto implemented is subject to great variation.

Consider Egypt. The country has a long-standing parliamentary tradition and strong preferences for democracy were demonstrated during the Arab spring movement. However, the country has also witnessed leadership by religious authorities and a military coup in the past decade, and corruption levels are rising. In a survey for the Arab Barometer project in 2016, the mode answer to the question "From 0 to 10, to what extent do you think democracy is appropriate for your country?" was 5, suggesting rather ambivalent attitudes. Germany, on the other hand, has been a democracy since the end of the second world war, despite looking back at a political history dominated by monarchists and fascist rulers. The Economist Intelligence Unit (2019) summarizes the current polarization between the two countries in the *Democracy Index*, which rates the regime in Germany as "fully democratic" and Egypt as "authoritarian".

Different notions of popular belief in what democracy actually comprises become evident if we compare answers to the questionnaire of the World Value Survey (Inglehart et al., 2014). In Germany, 71% of the people asked expressed that free elections are "an essential characteristic of democracy", compared to 45% in Egypt. At the same time, 17% of Egyptians strongly agreed that the interpretation of law by religious authorities is an essential part of democracy, compared to 2% of Germans. Surveys conducted by the Arab Barometer make the ambiguity in the Middle East visible: while most people believe democracy to be a desirable form of government, they are worried about the political outcomes it produces and whether it is fit for their particular societies (Robbins, 2015).

Economic experiments are a way to measure revealed preferences beyond the stated preferences elicited through surveys. Numerous experiments have established a phenomenon coined *democracy premium*: subjects cooperate more if an institution was brought about in a participatory way (Dannenberg and Gallier, 2020).

However, the laboratory studies of the economics discipline rely virtually exclusively on results from European and North-American studies of so-called Western, educated, industrialized, rich, and democratic (WEIRD) populations (Henrich, Heine, and Norenzayan, 2010). The research question of the present paper is if it increases cooperation to have a policy introduced through a representative, and how cultural and personal characteristics mitigate this effect. The contribution is thus twofold: quantify the democracy premium in the context of a representative democracy and investigate whether the findings are robust across cultures.

For the three stages of the experiment, subjects are matched into small groups. The first stage consists of a prisoners' dilemma, a game that represents in a simple way the conflict between maximizing individual benefits versus collective efficiency: both players could be made better off by cooperating, but each faces the incentive to deviate. In the second stage, subjects elect a representative, who can choose whether to modify the payoff structure for her group to a coordination game, which makes cooperation incentive-compatible, or to stick with the prisoners' dilemma. However, the representative's preference about the payoff modification is only considered with some probability. If it is not considered, then either the coordination game or the prisoners' dilemma is randomly assigned to each group for the second stage. This randomization technique developed by Dal Bó, Foster, and Putterman (2010) (hereinafter: DFP, 2010) avoids methodological problems of self-selection typically associated with voting in economic experiments. The exogenous intervention allows for a clean identification of the effect of the democratic choice. In the third stage, subjects play the game that was chosen for them in the second stage. The effects of the two ways to implement the payoff change – by the representative or by the computer – are obtained by observing cooperativeness in stage three.

The experiments were run in Cairo (Egypt) and Hamburg (Germany) for a direct comparison of two student populations in culturally different contexts: German students are predominantly Christian or atheist, while Islam is the state religion of Egypt along with a minority of Coptic Christians. Descending from several millennia of history, Egyptian culture is not solely defined by Islam. It is, however, worthwhile to investigate potential links between religiosity and democratic preferences, as the supposed incompatibility of Islam and democracy is an ongoing academic and public debate. By comparing undergraduate students from Germany and Egypt in the same experimental setting we get a clean estimate of the behavioral differences that arise from the different socialization of the two populations.

Regarding the results in Egypt, subjects are hesitant to introduce the cooperation-enhancing payoff modification and the way of implementation has very little influence on behavior. The estimate of the democracy premium is insignificant; if anything, a policy implementation by a representative decreases the willingness to cooperate. There is no support for hypotheses regarding a micro-level link between

democracy and religiosity in the data. In Germany, the democracy premium is substantial: Cooperation increases by more than thirty percent if the same policy is implemented by an elected representative and not by the computer. Especially subjects who initially did not prefer the cooperation-enhancing policy strongly increase their willingness to cooperate if it was brought about democratically.

The paper proceeds as follows. Section two reviews the relevant literature, section three presents the research design and hypotheses. The analysis including some robustness checks can be found in section 4 and section five discusses the results and finally concludes.

3.2 Related Literature

A large body of recent experimental literature investigates what is called the *democracy premium*: Subjects cooperate more with institutions that were democratically implemented (i.e. endogenous) as opposed to externally imposed ones (i.e. exogenous). A frequent result of these studies is the substantial size of the democracy premium. Tyran and Feld (2006), Sutter, Haigner, and Kocher (2010), Markussen, Putterman, and Tyran (2014), Gallier, Kesternich, and Sturm (2017), Dannenberg, Haita-Falah, and Zitzelsberger (2020), and Gallier (2020) show in laboratory experiments how voting on institutions fosters cooperation in social dilemma situations. A commonly encountered methodological issue in these experiments is that voting induces self-selection into treatments and thus biased estimates. DFP (2010) develop a new experimental technique, which is able to control for selection as well as information effects arising from the democratic policy choice to isolate the causal influence of the decision-making process. The design relies on randomization along with a strategy method to control for unobserved personal subject characteristics. DFP (2010) find evidence of a large democracy premium in setting modelling a direct democracy. The experiment presented here adapts the technique and applies it to representative democracies.

Henrich, Heine, and Norenzayan (2010) make the case for a more cautious handling of behavioral studies of WEIRD societies. Results purely obtained from such a small subset of the world population cannot be assumed to be representative of humanity as a whole and are of limited generalizability. This concern applies directly to the evidence on the democracy premium since procedural preferences and perceptions of legitimacy are inherently related to cultural norms, beliefs, and values. So far, laboratory experiments about the democracy premium were almost exclusively carried out in North-America and Europe.[1] One notworthy exception is Vollan et al. (2017), who investigate cooperation and authoritarian values in China. They find an

[1] There are some field experiments about the effects of participatory decision-making in non-WEIRD environments, predominantly in agricultural settings, e.g. Bardhan (2000), Cavalcanti, Schläpfer, and Schmid (2010), Olken (2010), Grossman and Baldassarri (2012), and Gallier, Langbein, and Vance (2018).

inverse relationship between authoritarian preferences and compliance with democratic decisions. The finding points in the direction of important cultural differences in terms of behavioral responses to decision-making procedures.

Even before Henrich, Heine, and Norenzayan (2010)'s appeal, there have been some – sporadic, but significant – efforts towards large-scale studies looking a cross-cultural differences in behavior. The workhorse games of those studies are the ultimatum game and public goods games .[2] Roth et al. (1991) and Henrich et al. (2001) compare ultimatum bargaining behavior across the world and find that offers are much larger than predicted by game theory and that differences between cultures are large. Chuah et al. (2007) and Chuah et al. (2009) find differences in offers made in ultimatum games between Malaysia and the UK to be explicable with answers to World Value Survey questions. For example, religiosity is negatively correlated with ultimatum offers. Gächter, Herrmann, and Thöni (2005), Henrich et al. (2006), and Herrmann, Thöni, and Gächter (2008) research norm enforcement and punishment behavior in public goods games and show, inter alia, that antisocial punishment – punishing cooperative individuals – is more prevalent in countries with weak institutions. Falk et al. (2018) conduct the Global Preference Survey to collect data on time and risk preferences, reciprocity, altruism, and trust, inter alia in relation to culture and religion, in 76 countries. Cohn et al. (2019) experimentally observe honesty and altruism in 40 different countries. Most of the studies presented here focus on the basal parameters of individual decisions like fairness and risk preferences. Looking at the democracy premium in different contexts implies investigating a much richer phenomenon, because the decisions in the coordination game are influenced by an interaction of personal preferences, social norms, and expectations.

Researching democratic and procedural preferences in a Muslim-majority country like Egypt provides the opportunity to gather evidence on the much contested (in-)compatibility between Islam and democracy. In the 1990s, Huntington (1996)'s *clash of cultures* provided a popular hypothesis of supposedly inherent differences between Western (Christian) and Eastern (Muslim) societies. Other – Western – authors sang the same tune, e.g. labeling Islam as a "grave threat to liberal practices" (Fukuyama, 1992, p.45). The rare occurrence of democracy in the Middle East was said to be explained by a general incompatibility between Islam – especially *Sharia*

[2] In the two-player ultimatum game as developed by Werner Güth, one player proposes a split of a given amount of money between her and the second player. If the second player accepts this proposal, the money is split as suggested. If the second player rejects, both earn nothing. Game theory suggests that the proposer offers the lowest amount possible, which is accepted by the responder. Experimental evidence is used to failing to confirm this equilibrium prediction (Güth, Schmittberger, and Schwarze, 1982; Güth and Kocher, 2014). In a public goods game, or voluntary contribution mechanism, subjects choose how to distribute their income between a private and a public account. All money put into the public account is multiplied by some factor and distributed equally between all group members. Depending on the multiplication factor, subjects have strong incentives to free-ride on others contributions to the public account, such that the equilibrium outcome is that every player exclusively pays into their private account. Again, experiments do not confirm the game theoretical predictions (Ledyard, 1994; Zelmer, 2003; Chaudhuri, 2011).

law – and democracy (Lipset, 1994; Kedourie, 1994). The ideas came back into fashion in the aftermath of the terrorist attacks of 9/11, when anti-Muslim sentiments spiked across Europe and the United States (Kumar, 2010). A different strand of literature argues against the incompatibility of Islam and democracy (Beinin and Stork, 1997; Entelis, 1997; Eickelman and Piscatori, 2018). Since the Quran incorporates the inherently democratic concepts of consultation, independent reasoning, consensus, and freedom of speech it cannot be anti-democratic per se (Esposito and Piscatori, 1991; Tayekh, 2001).

Tessler (2002, p.340) conjectures that since Islamic doctrine and institutions are used both to support as well as oppose democracy, the interpretation and implementation is crucial and the link between Islam and the absence of democracy not a causal one. A deeper investigation at the individual level is thus needed to explain the cross-country phenomenon that Muslim countries are seldom democratic (Hofmann, 2004). Tessler (2002) uses survey data of Sunni Muslims from Egypt, Palestine, Morocco, and Algeria between 1988 and 1996 to investigate the *political culture* – the opinions and norms present in society – as one pillar of democratic transition. He reports that of all countries measured, Egypt in 1988 has by far the highest share of respondents who find that democracy is not important, but only four years later, this share has dropped by over 40 percentage points.[3] It is found that Islamic religiosity is somewhat negatively correlated with support for democracy among Egyptian women. This gender difference as well as the time inconsistency in the results point to the sensitivity of the measures and to the fact that more research on the topic is needed to identify the relevant transmission channels. Collins and Owen (2012) find that religiosity among Muslims is negatively correlated with support for democracy, but not the religious affiliation with Islam per se. No empirical study was able to establish a distinct fault line between Christian and Muslim societies in terms of political values. A host of studies has investigated how attitudes towards Islam and democracy relate to one another and mostly failed to confirm the negative relationship that the clash of cultures would conjecture (Midlarsky, 1998; Ciftci, 2010; Ciftci, Wuthrich, and Shamaileh, 2019). The connection between Islam and political values is ambiguous and less strong than hypothesized by Huntington (1996). After all, Muslims and non-Muslims are seemingly not that different. However, the findings discussed here rely exclusively on survey data and are unable to demonstrate what behavioral consequences follow from the stated preferences. The present paper therefore combines questionnaire items about religion and political attitudes with an incentivised economic experiment, which is able to quantify cooperative behavior in mixed-motive games as well as reactions to more or less participatory decision-making procedures.

[3]The data is not a panel, but from two different surveys conducted in Cairo. While the samples' composition seems to be comparable, the questionnaire items differed, with the first asking about parliamentary democracy as the preferred political system and the latter one asking specifically about the desirability of open elections.

3.3 Research Design

3.3.1 Experimental Design

The experiment consists of three stages (see Figure 3.1) and is based on the standard prisoners' dilemma (Table 3.1). The prisoners' dilemma has a unique symmetric Nash equilibrium where both players choose action B (defect). Following DFP (2010) ten rounds of the prisoners' dilemma are played in the first stage with random rematching of pairs in every round within each group. The groups are made up of four players and remain together over the entire session. After each round, subjects' computer screen informs them about the actions of all players from their group.

FIGURE 3.1: Sequence of the Experiment

In the following vote stage, stage 2, subjects make two choices. First, they decide whether they want to change the payoff of their group to a coordination game for stage 3, or remain with their group in the prisoners' dilemma as known from stage 1. The coordination game has a Pareto-superior Nash equilibrium in mutually choosing A (cooperate). However, (B,B) remains as a Nash equilibrium also in the coordination game. Every subject privately states a preference about implementing the payoff modification or not. This decision will matter if she is elected as group representative (strategy method). Second, to elect the representative players privately announce another group member's ID. Subjects do not know their group members' modification preferences, but may base their choice on how cooperative a player was in stage 1. To this end, the computer screen shows again a table of all actions of the group members in stage 1, just as it did after each of the ten rounds. The player who is then named most often in a group is elected as representative via plurality rule. The computer randomly breaks any ties.

		Player 2	
		A	B
Player 1	A	$(50, 50)$	$(30, 60)$
	B	$(60, 30)$	$(40, 40)$

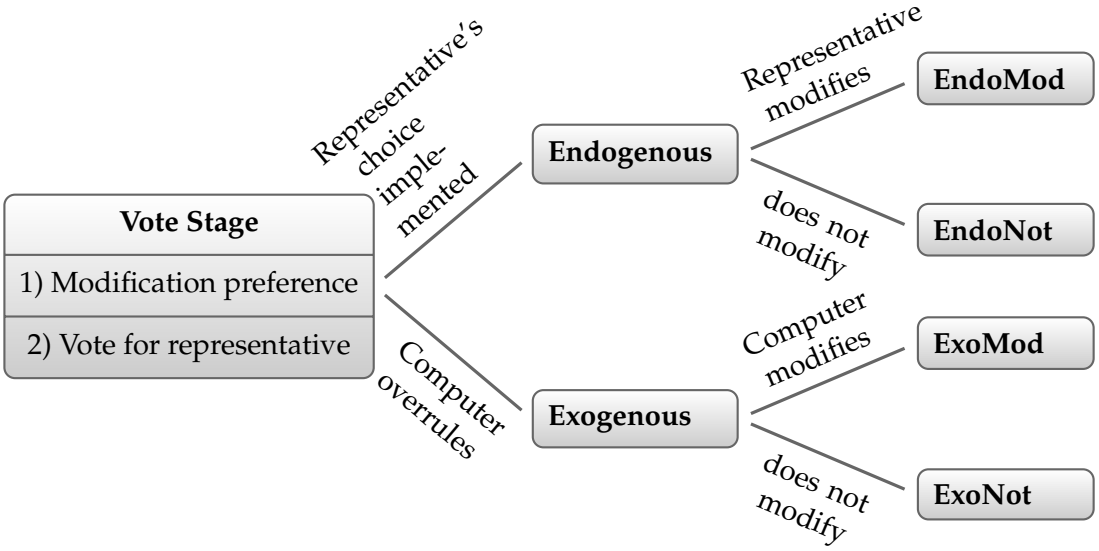
		Player 2	
		A	B
	A	$(50, 50)$	$(30, 48)$
	B	$(48, 30)$	$(40, 40)$

TABLE 3.1: Prisoners' Dilemma (left) and Coordination Game (right)

Analogous to DFP (2010), each representative's preferred game is implemented with probability $\frac{1}{2}$. If it is not implemented, the computer chooses either the prisoners' dilemma or coordination game for the group, again with probability $\frac{1}{2}$. As a consequence, there are four conditions under which subjects can play stage 3 (see

Figure 3.2): payoffs modified to a coordination game by the representative (EndoMod) or by the computer (ExoMod), and the unmodified prisoners' dilemma game either chosen by the representative (EndoNot) or exogenous (ExoNot).

FIGURE 3.2: Vote Stage Outcomes (adapted from DFP, 2010)



The twofold random intervention makes it possible to compare groups whose representatives decided in the same way but who ended up in different conditions. This is crucial to control for self-selection. Assuming there are unobservable player characteristics that increase both the preference for the coordination game as well as the willingness to cooperate, the treatment assignment is non-random whenever we let subjects choose their own payoff structure. However, as DFP (2010) show, once the analysis conditions on an individual's original preference and the payoff structure, correlated unobservable characteristics are controlled for.

Immediately after the vote stage, subjects are informed about the player ID and game choice of the representative, whether the choice was considered, and the game the group will ultimately play in the last stage. As a consequence the experimental design not only controls for payoff modification preferences by usage of the strategy method, but also holds information constant across outcomes.[4] The respective game is then played for another ten rounds in Stage 3.

3.3.2 Hypotheses

The design allows observations between subjects (by comparing behavior between the different vote stage outcomes) and within subjects (by comparing subject behavior before and after the vote stage). The democracy premium is hard to reconcile with even behavioral game theory, since the way in which a situation was reached

[4]If the representative's choice of game was revealed only in the endogenous condition, as is the case in DFP (2010)'s main treatment, subjects are able to update their beliefs about their group members in a way that the subjects in the exogenous case could not.

should not influence behavior as long as the information and actions available remain the same. Therefore, the following hypotheses are derived from previous empirical insights. First, DFP (2010) find that subjects who cooperate more in the prisoners' dilemma are indeed more likely to prefer the coordination game.

Hypothesis 8 *A subject who is more cooperative in stage 1 is more likely to have a preference for the coordination game.*

Second, since the subjects do not observe the other players' payoff preferences at the time when they elect the representative, they have to base their voting decision exclusively on the others' behavior in the first stage. Hamman, Weber, and Woon (2011) find in a comparable setting that groups tend to elect prosocial representatives if those get to make decisions on issues that require the group to cooperate.

Hypothesis 9 *Cooperative players in stage 1 are more likely to become representatives.*

Third, the central hypothesis regarding the existence of the democracy premium requires subjects to cooperate more in the coordination game in stage 3 if it was chosen by their representative as opposed to the computer (EndoMod versus ExoMod). To be unbiased by self-selection this further requires to control for the individual preference (DFP, 2010). If the democracy premium is indeed a universal phenomenon, Hypothesis 3 should be confirmed in Egypt as well as in Germany.

Hypothesis 10 *Cooperation rates in the coordination game are higher under endogenous modification compared to exogenous modification after controlling for individual game preferences.*

Fifth, the origins of the democracy premium as described in Hypothesis 4 are still obscure to the research community. This paper proposes two related transmission channels in the context of a representative democracy: signalling and obedience towards authority. To look for a signalling effect from the representative's choice, we can compare the difference-in-differences between stage 1 and 3 if the representative chose the coordination game or not. To respect the ceteris paribus assumption this can only be done in the unmodified conditions, as otherwise the payoff structure would change between the stages. If the representative's game preference matters for behavior despite having been overruled that would be strong evidence in favor of the representative's ability to create relevant focal points for her constituents.

Hypothesis 11 *Cooperation rates of subjects in ExoNot are higher in stage two compared to stage one if the representative chose the coordination game but was overruled.*

Sixth, the setting in Egypt allows for the analysis of experimental play in combination with personal characteristics outside of the Western and Christian sphere that is the predominant experimental paradigm in the literature on the democracy premium. Vollan et al. (2017) find for a Chinese subject pool that the adherence

to authoritarian values decreases the democracy premium significantly. Collins and Owen (2012) find a negative relationship between religiosity and support for democracy among Muslim subjects. Both conjectures can be tested with the experiment at hand.

Hypothesis 12 *Strong religiosity or obedience towards authority decrease the democracy premium.*

3.4 Analysis

3.4.1 Data Sources and Procedures

The experiment was conducted with the zTree software (Fischbacher, 2007). In Germany, 140 students were recruited via hroot (Bock, Baetge, and Nicklisch, 2014) at the University of Hamburg between 2016 and 2018.The same study was replicated in Cairo at the British University in Egypt with 180 students in 2019. All Egyptian participants were students from the faculty of economics, business administration, and political science of the British University in Egypt, which were recruited via email invitation to sign up for experimental sessions. The lab experiment in Cairo is complemented by select questionnaire items from the World Value Survey and Arab Barometer, which cover religious and authoritarian values (see Appendix Section 3.6). Sessions were held at the computer pool of the main library at the British University in Egypt and in the computer lab of the department of economics in Hamburg. In both countries the participants were randomly allocated to computers upon arrival to the session. Instructions were read aloud and complemented by control questions before the experiment. The instructions in Germany were written in German. In Cairo, the instructions were translated into English, which is the official language for teaching and administration at the British University. Both subject pools answered an unpaid socio-economic questionnaire at the end of the session, which included the aforementioned items on religiosity in Egypt. Payment was made individually and in cash directly to the subjects after the session. Subjects' anonymity was protected in all cases to make sure that decisions in the experiment cannot be linked to personal data.

3.4.2 Main Analysis

The following sections examine the experimental data in the light of the hypotheses, followed by robustness tests. Unless indicated otherwise, p-values are derived from non-parametric statistical tests, i.e. Wilcoxon rank-sum test for within- and Mann-Whitney-U-test for between-subjects comparisons. Table 3.2 shows the summary statistics. Egyptian subjects were on average a little younger and a little more male than Germans. The average payoff in Cairo was 85 Egyptian pounds, which at the

time of the study corresponded to 5.3 USD. Subjects in Hamburg earned 9 Euros on average (9.9 USD) and all sessions lasted for less than one hour.[5]

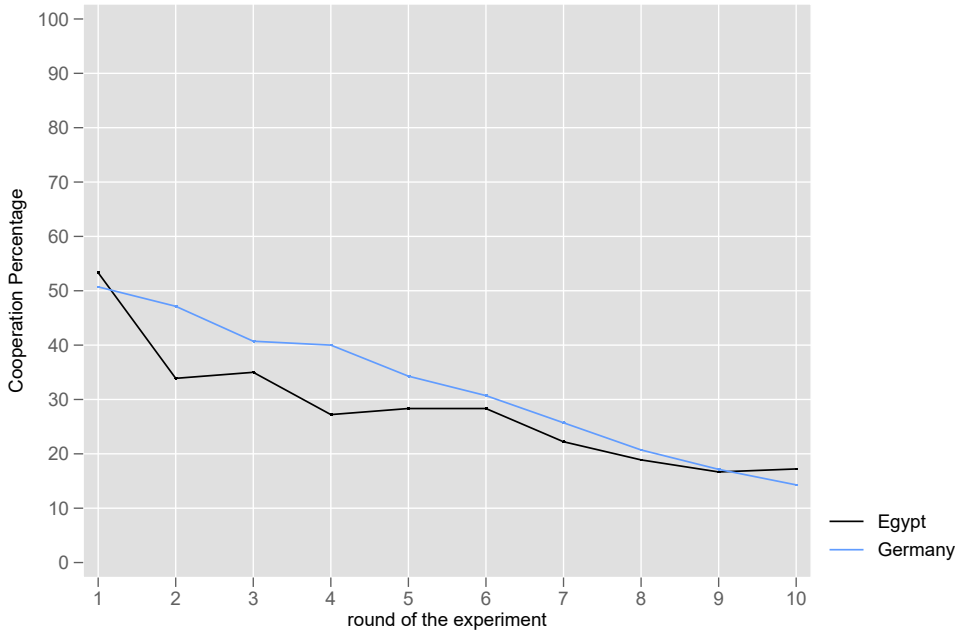
TABLE 3.2: Summary Statistics

	Egypt	Germany
Sample Size	180	140
	Mean	
Age	20	25
Payout	85 EGP	9 EUR
	Share	
Female	48 %	60 %
Muslim	88 %	n.a.
Christian	11 %	n.a.

Stage 1: Prisoners' Dilemma

Cooperation in the first ten rounds of the prisoners' dilemma averaged at 28 percent in Egypt and 32 percent in Germany (p = 0.01). Figure 3.3 shows the commonly observed pattern of decreasing cooperation rates over time (Cooper et al., 1996). In Germany, women are on average 17 percent more cooperative than men ($p < 0.001$). Neither gender nor religion are correlated with cooperation in Egypt.

FIGURE 3.3: Cooperation Rates Part 1 – Prisoners Dilemma



[5]The payoffs for both countries were calibrated to reflect the hourly wage of a student assistant.

Stage 2: Vote Stage

Payoff Modification When asked about their preference to modify payoffs, in Egypt a minority of subjects (43 %) preferred the potentially cooperation-increasing coordination game. In Germany, 62 % chose to modify payoffs. When regressing the preference for the coordination game on individual player characteristics the players that cooperated more in stage 1 are more likely to prefer the coordination game, lending support to Hypothesis 1 (Table 3.3). Ceteris paribus, a subject that cooperates in all rounds of stage 1 is 41 percentage points more likely to modify the payoffs. The effect is particularly strong for the Egyptian subsample (column 2), and not significant for the German subject pool. The explanatory variable partners' cooperation measures the actions taken by a subject's opponents. Higher values of partners' cooperation decrease the likelihood to prefer the coordination game, potentially because subjects facing already cooperative opponents see less necessity to modify the game. The effect is less strong than own cooperation and not significant for the subsamples. There is furthermore a highly significant gender effect: women are less likely to choose the coordination game, which is driven entirely by the Egyptian subjects. There is no effect of religion.

Result 9 *Cooperative individuals have a higher preference for the coordination game.*

Election of the Representative Hypothesis 2 states that more cooperative players are more likely to be elected as representatives. Table 3.4 shows that indeed being more cooperative increases the likelihood of being elected. The effect stems from the German sample, where cooperating in all rounds of the first stage makes someone 22 percentage points more likely to become representative as opposed to someone who never cooperates (column 3). The preference for payoff modification, which is not observable at the time of the election, is not significantly correlated to becoming representative.

Result 10 *Elected representatives are not more cooperative than other players in Egypt. In Germany, being more cooperative increases the likelihood of being elected.*

Table 3.5 shows the outcome of the vote stage and how many players ended up in the four different conditions for stage 3. Due to the hesitation of Egyptian subjects to modify payoffs, in the endogenous outcome subjects are significantly more likely to be in the prisoners' dilemma ($p<0.01$). In Germany relatively more subjects switched to the coordination game in the endogenous condition ($p<0.001$).

Stage 3: Democracy Premium

Figure 3.4 shows that in both countries cooperation was significantly higher under the modified payoffs: given payoff modification subjects mostly able to coordinate on the more efficient equilibrium. For both games the respective cooperation rates

TABLE 3.3: Determinants of Preference for the Coordination Game

Dependent variable: Modification Preference			
	(1)	(2)	(3)
	all	Egypt	Germany
Own cooperation	0.041***	0.045**	0.038
	(0.013)	(0.017)	(0.023)
Partners'	-0.026*	-0.020	-0.026
cooperation	(0.015)	(0.022)	(0.025)
Female	-0.140**	-0.190**	-0.095
	(0.055)	(0.074)	(0.081)
Egypt	-0.200***		
	(0.057)		
Muslim		0.150	
		(0.097)	
Constant	0.659***	0.321**	0.643***
	(0.068)	(0.123)	(0.083)
N	320	180	140
R^2	0.091	0.082	0.039

Note: Linear probability model with the dependent variable equal to 1 if a player preferred the modified payoffs. *Own cooperation* and *partners' cooperation* are count variables between 0 and 10 for each round of cooperation in stage 1. Standard errors (clustered at group level) in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3.4: Probability of Winning Election

Dependent variable: Elected as Representative			
	(1) all	(2) Egypt	(3) Germany
Own cooperation	0.016*	0.009	0.022*
	(0.010)	(0.016)	(0.013)
Modification	-0.087	-0.132	0.014
	(0.058)	(0.085)	(0.072)
Constant	0.375***	0.448***	0.249***
	(0.051)	(0.063)	(0.077)
N	320	180	140
R^2	0.014	0.018	0.023

Note: Linear probability model with the dependent variable equal to 1 if a player was elected as representative (before tie-breaking). *Own cooperation* is a count variable between 0 and 10 for each round of cooperation in stage 1. *Modification* is a dummy variable equal to 1 if a subject preferred the modified payoffs. Standard errors (clustered at group level) in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3.5: Number of Subjects per Vote Stage Outcome

	Prisoners' Dilemma	Coordination Game	Total
	Egypt		
Implementation:			
Endogenous	56	36	92
Exogenous	36	52	88
Total	92	88	180
	Germany		
Implementation:			
Endogenous	24	48	72
Exogenous	36	32	68
Total	60	80	140

were lower in Egypt compared to Germany (p=0.05 for the coordination game and p=0.06 for the prisoners' dilemma). The differences in behavior between the two games imply that the payoff change itself was effective in increasing cooperativeness. In the following we will disaggregate the results further to gather evidence also about the effect of the implementation mechanism.

FIGURE 3.4: Cooperation rates by Game and Country in Part 2

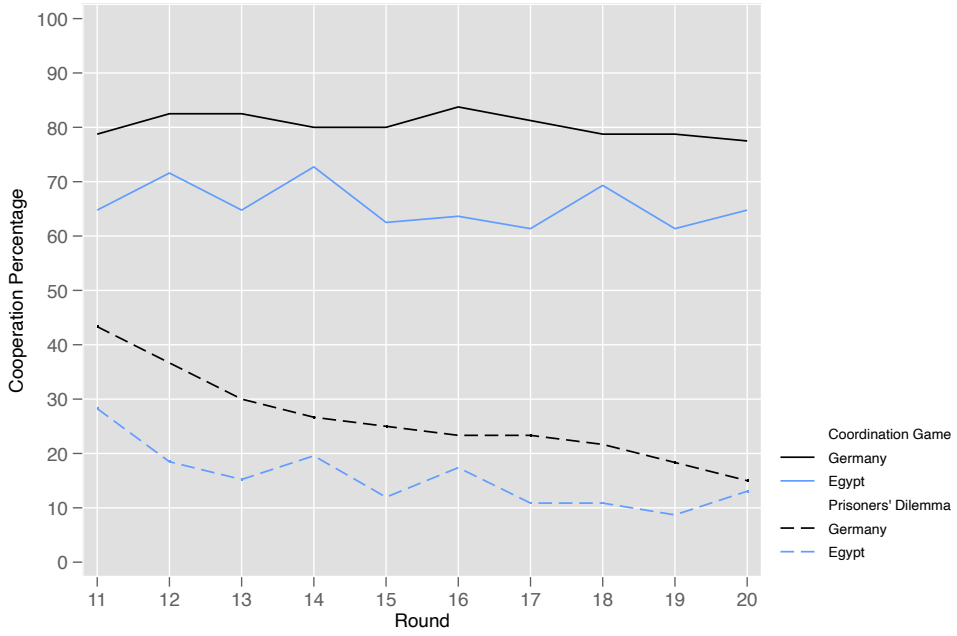


Table 3.6 gives the number of observations and cooperation rates directly after the vote stage separated by institutional preference and vote stage outcome in Egypt and Germany respectively. The topmost panels break down the subject numbers from Table 3.5 by individual modification preference. Focusing on the results from Cairo first, we see that selection effects are indeed relevant for the estimation of the democracy effect: there are more than twice as many yes-voters in EndoMod compared to ExoMod (69 versus 33 percent, $p < 0.001$). Given the previous insight that cooperative players are more likely to vote yes on the payoff modification, the imbalance in group composition would be the source of a veritable bias in a naive estimation of the democracy effect that does not control for self-selection. In Germany there is no statistical difference between the share of subjects preferring the coordination game in the endogenously and exogenously modified conditions (71 and 69 percent, $p = 0.84$).

In Egypt, in the first round after the vote stage cooperation is higher if the coordination game was *not* introduced by the representative (ExoMod vs. EndoMod). The main difference is between the subsets of yes-voters – those in favor of the modification – between the two outcomes: even though their representative was not considered, yes-voters in ExoMod have on average a 24 percentage points higher

cooperation rate ($p = 0.08$). No-voters' behavior is statistically indistinguishable between endogenous and exogenous modification ($p = 0.88$). In Germany, the opposite holds true and cooperation is higher if the policy is democratically introduced, an effect mainly driven by the no-voters, who cooperate significantly more in EndoMod compared to ExoMod ($p = 0.08$).

TABLE 3.6: Individual Behavior After the Vote Stage

Egypt					
	Endogenous Condition		*Exogenous Condition*		
Prefer to modify	EndoMod	EndoNot	ExoMod	ExoNot	Total
Number of subjects in each outcome by preference:					
No	11	36	35	20	102
Yes	25	20	17	16	78
Total	36	56	52	36	180
Cooperation rate in Round 11 (percent):					
No	54.5	30.6	57.1	25.0	
Yes	64.0	40.0	88.2	12.5	
Average	61.1	33.9	67.3	19.4	

Germany					
	Endogenous Condition		*Exogenous Condition*		
Prefer to modify	EndoMod	EndoNot	ExoMod	ExoNot	Total
Number of subjects in each outcome by preference:					
No	14	13	10	15	52
Yes	34	11	22	21	88
Total	48	24	32	36	140
Cooperation rate in Round 11 (percent):					
No	64.3	23.1	40.0	46.7	
Yes	94.1	27.3	81.8	61.9	
Average	85.4	25.0	68.8	55.6	

Using the cooperation rates and subject numbers of Round 11 we can determine the size of the *democracy premium*: the increase in cooperation that cannot be explained by the different payoff structures of the two games or group composition. Following DFP (2010) the decomposition of policy effects can be calculated in the form of weighted differences in cooperation rates in Round 11, as they are given in Table 3.6 (see Table 3.7 for an overview and Appendix Section 3.6 for details of the calculation). The weighting by modification preferences accounts for differences in group composition: as modification preference and cooperation are positively correlated, a group with more yes-voters can be expected to have higher cooperation rates. Therefore, the cooperation rates in Table 3.6 are separated by modification preference and enter the estimation weighted by these shares. The total policy effect

quantifies the overall impact of introducing the coordination game by the representative. It corresponds to the unweighted difference in cooperation between endogenous payoff modification and non-modification (EndoMod and EndoNot). It is more than twice as large in Germany (60 percent) than in Egypt (27 percent): the Egyptian subjects are less sensitive to the institutional change, i.e. they are less able to switch to the more efficient equilibrium once the coordination game is in place.

The total policy effect can be decomposed into the selection effect, the exogenous treatment effect, and the democracy premium. The selection effect accounts for the fact that the total policy effect overestimates the effect of the endogenous payoff modification, since there are fewer yes-voters in EndoNot than in EndoMod. The selection effect estimate gives the increase in cooperation that would be observed in EndoNot if the group composition was the same as in EndoMod. It amounts to 3 percent in Egypt and 1 percent in Germany. The exogenous treatment effect measures the pure effect of switching from prisoners' dilemma to the coordination game, i.e. the difference between ExoMod and ExoNot. The exogenous treatment effect is more than five times larger in Egypt than in Germany: 62 versus 12 percent. The residual part of the total policy effect, which cannot be accounted for by selection or the exogenous treatment effect, is the democracy premium. While the selection effects are in either case rather negligible, the democracy premium is substantial in Germany: It amounts to 47 percentage points, and thus accounts for more than two-thirds of the total policy effect. The effect in Egypt is of similar magnitude, but *negative*: subjects cooperate less if the policy was introduced by a representative compared to the randomization device. The p-values of tests for the significance of differences in cooperation rates between vote stage outcomes are summarized in Table 3.8. We see that the payoff difference between the games is all that matters in Egypt, not the implementation mechanism. In Germany, all four outcomes cause significantly different behavior.

TABLE 3.7: Policy Effects in Comparison

	Egypt	Germany
Total policy effect	27.1***	60.4***
Selection effect	3.2	1.1
Exogenous treatment effect	62.4***	12.1
Democracy premium	-38.4**	47.2**

Note: Policy effects are obtained from weighted differences in cooperation rates in Round 11.

The findings are corroborated by the regression results reported in Table 3.9, which explains the probability of cooperating in Stage 3. Using the ExoMod condition as baseline, EndoMod is significantly negative in Egypt in interaction with the individual modification preference. Thus, subjects cooperate less if they receive their preferred payoff structure via the representative. Unsurprisingly, both unmodified conditions are negatively correlated with cooperation. In Germany, the influence of

TABLE 3.8: Individual Cooperation By Vote Outcome

p-values for differences between cooperation rates

	Egypt	Germany
Variation in Payoffs		
EndoNot vs. EndoMod	0.00***	0.00***
ExoNot vs.ExoMod	0.00***	0.00***
Variation in Implementation		
EndoMod vs.ExoMod	0.62	0.00***
EndoNot vs.ExoNot	0.16	0.00***

Note: p-values are obtained from Mann-Whitney tests using cooperation rates of all rounds of Stage 3.

EndoMod compared to ExoMod is positive: subjects cooperate more if the coordination game is introduced by the representative. Most interesting is the interaction term between EndoMod and modification. The endogenous modification has the most positive effect on those who initially did not prefer modification.

Result 11 *Cooperation is higher in Germany if the policy is introduced by a representative and the size of this democracy premium is substantial. There is no democracy premium in Egypt.*

Figures 3.5 and 3.6 show individual cooperation for all conditions separated by individual voting behavior for all rounds of the experiment. We see that in both countries the subjects who initially did not vote for modification react most strongly and positively to the endogenous modification. Such behavior is in opposition to DFP (2010), where the democracy premium in the direct democracy setting is driven by the yes-voters. The mechanics of democratic legitimization thus seem to differ between direct and indirect democracy as well as between political cultures. The elected representative does not only bring the democratic legitimacy, but is also an authoritative figure and a strong focal point. I will therefore use the first of the following robustness tests to investigate further how attitudes towards authority and signalling effects shape responses to the voting procedure.

FIGURE 3.5: Cooperation Rates by Game and Preference in Egypt in Stage 3

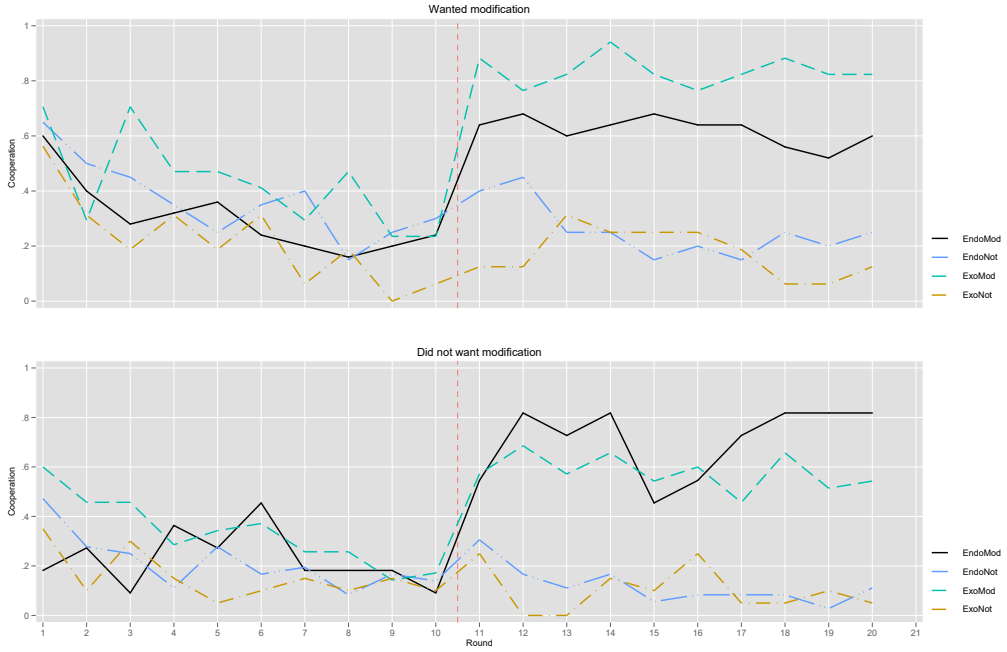


FIGURE 3.6: Cooperation Rates by Game and Preference in Germany in Stage 3

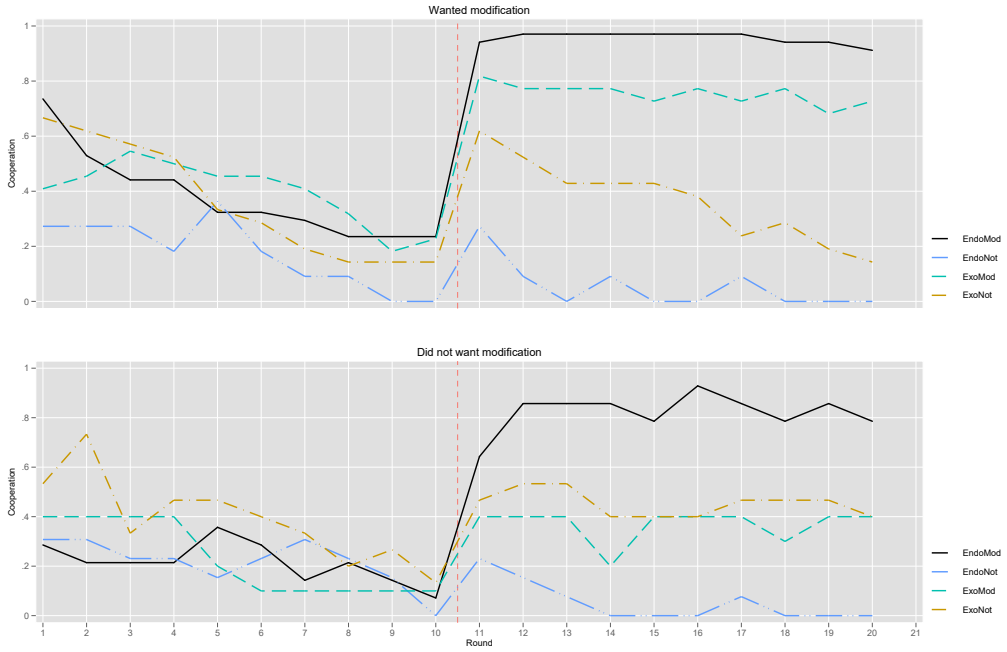


TABLE 3.9: OLS Regression - Cooperation by Vote Stage Outcome

Dependent variable: Cooperation Probability in Stage 3

	(1) Egypt	(2) Germany	(3) Egypt	(4) Germany
EndoMod	-0.0581	0.280**	0.129	0.451***
	(0.109)	(0.134)	(0.0993)	(0.158)
ExoNot	-0.544***	-0.221	-0.480***	0.0833
	(0.0927)	(0.169)	(0.0959)	(0.215)
EndoNot	-0.499***	-0.557***	-0.461***	-0.316**
	(0.0876)	(0.130)	(0.0937)	(0.138)
Modification	0.114***	0.102*	0.255***	0.385***
	(0.0358)	(0.0548)	(0.0706)	(0.0926)
Round	-0.0101**	-0.0130**	-0.0101**	-0.0130**
	(0.00489)	(0.00595)	(0.00490)	(0.00595)
EndoMod*Modification			-0.344***	-0.250**
			(0.113)	(0.117)
ExoNot*Modification			-0.180**	-0.471***
			(0.0795)	(0.158)
EndoNot*Modification			-0.120	-0.384***
			(0.0833)	(0.0980)
Constant	0.783***	0.766***	0.737***	0.572***
	(0.0876)	(0.146)	(0.0879)	(0.148)
N	1800	1400	1800	1400
R^2	0.279	0.403	0.292	0.430
adj. R^2	0.277	0.401	0.289	0.427

Note: Standard errors (clustered at group level) in parentheses. Regressors in (3) and (4) are binary interaction terms of vote stage result and modification preference (suffix "y" pro and "n" against modification). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.4.3 Robustness Tests

This sections checks for the robustness of the results obtained from the individual analysis and digs deeper into specific aspects of the experimental design to elaborate on Hypotheses 4 and 5. Furthermore, an alternative identification strategy is presented.

Relevance of the Representative

First, let us shed more light on the role of the representative. If the legitimacy of having been elected by the group initiates behavioral change it could be the case that the signal of a representative who wants to modify payoffs works independent of the actual payoff change. Comparing cooperation rates between groups that were exogenously assigned the prisoners' dilemma, does it matter which game the overruled representative chose? It does. However, the effect of choosing the coordination game is rather *negative*: In Germany, subjects in this condition with a representative preferring the prisoners' dilemma cooperated on average 7.8 times in the 10 rounds of stage 3. Subjects with a representative that chose the coordination game but was overruled cooperated on average only 2.2 times out of 10 ($p < 0.001$). In Egypt, the difference is virtually non-existent: 2.1 compared to 1.7 rounds ($p = 0.77$). Following Hypothesis 4, it is also tested whether cooperation rates of subjects in ExoNot are higher in stage 3 compared to stage 1 if the representative chose the coordination game but was overruled. It is neither the case in Germany ($p = 0.14$) nor in Egypt ($p = 0.40$).

Result 12 *Having a representative who wants to modify payoffs does not increase cooperation rates under exogenously unmodified payoffs.*

Religiosity, Obedience, and Trust in Egypt

Using the more extensive questionnaire from Egypt, we can deduct whether and how personal attitudes are related to behavior in this particular population. Following Hypothesis 5, the section focuses on attitudes towards authority as well as religiosity (see appendix section 3.6 for a discussion of the entire questionnaire). First, consider the prisoners' dilemma in stage 1: The first column of table 3.10 shows that whether someone is Muslim or Christian does not matter for cooperativeness per se, but being especially devout does. The subjects of both denominations who pray and fast most rigorously are seven percent less likely to cooperate in a given round of the prisoners' dilemma. Second, let us also consider the preference for the coordination game (Column (2), Table 3.10). The only significant effect here is that women are 17 percent less likely to modify payoffs, but religion plays no role. But what we are most interested in here is the relationship between a subject and their representative. This is why, third, we look at the behavior in the second part. *Muslim* and *devout* are not significant control variables when regressing cooperation on the

four vote outcomes (regression table not reported). A further split sample analysis between religious affiliations (Table 3.11) shows that endogenous modification is not a significant predictor of cooperative behavior in stage 3 for Muslims, Christians or the subset of subjects that are the most religious (*devout*). The effects of all variables in the model are very similar across the three estimates. The probability to cooperate in a given period of stage 3 only depends on the game and the mechanism, but not on a subject's religion: Christians and Muslims in Egypt are equally unresponsive to the endogenous modification. Overall, religiosity is not a relevant explanatory factor for cooperation or democratic preferences in the experiment.

To test the second part of Hypothesis 5, let us focus on obedience towards authority (see also Appendix Section 3.6). Are subjects with an authoritarian disposition more or less responsive to the representative democracy? On the one hand, they should value participatory procedures less, but on the other hand they could be more likely to follow their representatives. Table 3.12 shows that these subjects are generally a little more cooperative than others (column 1), but that they react negatively to all vote outcomes when compared to exonot. They are significantly less likely to cooperate if the payoff modification was introduced by their representative (column 2).

Incorporating trust into the equation paints an entirely different picture: Believing that "most people can be trusted" does not make subjects more cooperative per se. But it makes them react more cooperatively to the representative's modification as opposed to the random modification. But because the share of subjects agreeing to that statement in Cairo is small, they fail to make an overall impact on average behavior (see also Section 3.6). Self-reported risk aversion levels are low in the sample and possess no explanatory power regarding the effect of democracy.

Result 13 *Religion is not a relevant factor for responsiveness to representative democracy. Obedience towards authority makes subjects react negatively to the democratic implementation. Higher trust in others yields a positive reaction to representative democracy.*

Group Level Analysis

An alternative identification strategy to control for individual modification preferences is to consider groups with identical preference structures. Analogous to DFP(2010, p.2212), the variable *voteshare* indicates how many subjects per group wanted to modify the payoff matrix in case they were elected as representative. Thus, there is only one observation considered per group, which is the average cooperation rate of its four members. Groups that share the same proportion of yes-voters can then be compared in the different vote stage outcomes. The focus is on groups with two votes for modification. For this value of *voteshare* all four possible outcomes were realised in the experiment making it possible to compare groups that have identical preferences but ended up under endogenous or exogenous modification (see Tables 3.13 and 3.14). Given that fewer subjects preferred the coordination

game in Egypt, the average value of *voteshare* is 1.7 in Egypt and 2.5 in Germany (p-value < 0.001). Comparing the average cooperation rates in the coordination game between the two countries, we see that the difference between EndoMod and ExoMod in Germany is 28.3 percentage points (p-value = 0.08), whereas in Egypt it is -1.5 percentage points (p-value = 0.71). This corroborates the results obtained from the individual analysis.

When holding group composition constant and comparing only groups that had an equal number of votes in favor of and against modification (*voteshare* = 2), for the German sample we find again a higher cooperation rate in EndoMod compared to ExoMod. The average in stage two is more than twice as large (p-value = 0.08). In Egypt, the effect runs in the opposite direction: groups with *voteshare* equal to 2 cooperated in 84 percent of the cases in ExoMod, but only 44 percent in EndoMod (p-value = 0.08). To summarize, the pattern from the individual analysis shows up even more prominently on the group level: There is a robust democracy premium in Germany, but not in Egypt.

3.5 Discussion

The effect of democracy in Germany is large, positive, and significant: the indirect democratic implementation leads to increased cooperative behavior. The effect is even bigger than in previous studies on direct democracies. Especially the subjects who initially opposed the institutional change cooperate highly if – and only if – it was brought about by an elected representative. In Egypt, there is very little difference between the representative and the computer choosing a policy, and the democracy premium is negative: if anything, Egyptian subjects cooperate less if a representative chose the policy. Moreover, in Germany the preference for the potentially more efficient coordination game is more prevalent. The majority of Egyptian subjects actually prefers the prisoners' dilemma over the coordination game. In both countries the more cooperative subjects are more likely to prefer playing the coordination game. A further insight from the first robustness test is that the representative's actions have to be complemented by an actual institutional change and intentions are not effective on their own. Thus, even though the decision-making process has behavioral consequences that go beyond the decision's outcome, these are not independent of the outcome.

What drives these results? The extensive additional questionnaire from Egypt allows to test for correlations between personal attitudes and behavior in the experiment. There is no significant relationship between adherence to Islam and democratic preferences, neither on the extensive nor on the intensive margin. This finding on the micro-level nexus, or lack thereof, between Islam and political values speaks directly to a lively debate in the political economics literature. The insight from this experiment is the following: subjects in the Muslim-majority country are less responsive to democracy, but Islam has no causal impact. Instead, obedience

to authority and a general propensity to trust others predict cooperative behavior well: being more trusting makes subjects more likely to coordinate on the payoff-dominant equilibrium of mutual cooperation. Whereas an adherence to traditional values makes the risk-dominant defection equilibrium more focal. We can deduct that religion and culture are indeed distinct concepts. The Egyptian subjects are of different religions and display varying degrees of religiosity, but having been raised in Egypt they share the same cultural background. The experimental findings so far point towards culture as the more important explanation of different reactions to the representative democracy in the experiment. Similar behavioral patterns would thus be expected not in every other Islamic country, but only in those that exhibit comparable values in society.

Other potentially relevant factors besides culture, which captures stable differences in attitudes and conventions between populations, are more imminent political disparities. On the Corruption Perceptions Index, Germany's score is more than twice as high than that of Egypt (Transparency International, 2020). Widespread (perceived) corruption in everyday life makes it much less attractive for Egyptians to follow recommendations of authorities and put trust in institutional procedures, which is reflected in the questionnaire answers on leadership: More than two thirds of the Egyptian subjects do not agree that "our leaders know what is best for us". Being exposed to an environment of unstable and untrustworthy authorities results in reluctance to rely on official procedures to produce efficient outcomes. Together with the overall low trust level it is evident why Egyptian subjects would not have a pronounced desire to switch to the coordination game if they expect to defect anyways.

Another potential explanation for the democracy premium itself as well as for the cultural differences is the distinction between the input and output legitimacy of a decision (Scharpf, 1999). Input legitimacy captures the degree to which political processes are responsive to the citizens preferences, while output legitimacy is concerned with the effectiveness of the chosen policies. Holding the output, i.e. the payoff structure, constant in the experiment, it is the input that varies in legitimacy between the democratic and the random implementation. It seems that German subjects are much more concerned with the input legitimacy than their Egyptian counterparts. Anecdotal evidence from Egypt suggests that citizens are predominantly interested in the output of a process: the first democratic elections after the Arab spring revolution brought the Muslim Brotherhood to power. Among those who were unhappy with their ruling it is common to blame democracy for producing the undesirable outcome and therefore being an unfit system for the Egyptian society (Tessler, 2002, p.85). Attaching value to input legitimacy independent of the outcome seems like a necessary condition for a democracy premium to arise. These different perceptions of legitimacy are undoubtedly a component of the political culture in a society.

The results reported in this paper are simple snapshots from two very different

subject pools operating under very different circumstances. But the results speak directly to the community of experimental researchers, who must take personal as well as cultural factors into account when trying to generalize their findings. The results have policy implications as well: what works in one context, might not work in another. Legal transplants and recommendations of foreign experts can fail to achieve their desired outcome even if the policies themselves match the communities' preferences.

3.6 Conclusion

The experiment measures the influence that democratic decision-making processes have on the effectiveness of a policy. Economic theory would not predict an observable impact of the decision-making process on behavior independent of the procedure's material outcome. Previous empirical studies found evidence of a *democracy premium*. However, the exact mechanisms and universality of the phenomenon remain obscure. The present paper's contribution to this literature is twofold: by extending the direct democratic procedures exclusively used beforehand to a representative democracy, and by comparing two culturally heterogeneous subject samples. To this end, the experiment modeling a representative democracy was conducted in Egypt and in Germany.

The design follows a mechanism introduced by Dal Bó, Foster, and Putterman (2010) that controls for self-selection effects, which would otherwise bias the analysis: more cooperative players are more likely to vote for cooperation-enhancing policies. As a consequence, if given the choice between two policies those types are overrepresented in the groups that implement the cooperation-enhancing institutions democratically and a naive estimator would produce overly optimistic results regarding the effect of democracy on cooperation. The design that avoids such a bias is as follows: Small groups elect a representative, who is presented with the possibility of changing their groups's payoff structure from a prisoners' dilemma to a coordination game that makes cooperation incentive-compatible. The modification preference of each player is used as a proxy for unobservable personal characteristics that influence cooperative behavior. It is tested whether the institutional change has the same influence if it is implemented by the group representative compared to random implementation. The representative's choice of game is considered for the group with a 50-percent probability. If it is not considered the computer chooses one of the two games with equal probability. The randomization allows for a comparison of subjects with the same preferences, information, and institution who only differ in the way the institution was implemented: by the representative or by the computer. The results allow an analysis of the effects of the decision-making procedure beyond its outcome. Controlling for an individual's policy preference allows the identification of the democracy premium without self-selection or information effects as potential biases.

It is found that the democratic implementation has a positive effect on cooperation only in Germany. Egyptian subjects have much lower preferences for even introducing the cooperation-enhancing institution and do not increase cooperation if it was brought about by an elected representative. Several robustness checks corroborate the results. Cooperation rates are not significantly related to religiosity or affiliation with Islam. But authoritarian attitudes and distrust towards other members of society decreases cooperativeness among Egyptians.

What is needed from future work on the topic is a synthesis of the findings related to the democracy premium and a more precise investigation of the channels through which participatory decision-making translates into higher willingness to cooperate (or not). Candidate theories can be drawn from social psychology and decision theory. Procedural utility allows actors to have preferences over how outcomes are reached, but cannot explain how the procedural preferences translate into behavior (Frey, Benz, and Stutzer, 2004). Group identity has also been considered a relevant factor in explaining the democracy premium (Akerlof and Kranton, 2010). Self-determination theory stresses the relevance of individual autonomy for intrinsic motivation (Ryan and Deci, 2000). It is feasible that being considered by the randomization device in the experiment increases one's motivation to cooperate. However, self-determination was shown to be a universal phenomenon and thus cannot serve to explain different democratic preferences between cultures (Church et al., 2013). Future research needs to provide a synthesis of the results from various contexts to enable the pinning down and specific testing of specific transmission channels between and within cultures. This is not only relevant to the scientific field itself, but provides insights to help nurture strained democratic procedures globally.

Bibliography

Akerlof, George A and Rachel Kranton (2010). "Identity economics". In: *The Economists' Voice* 7.2.

Al Jazeera (2019). "Lebanon protests: Thousands demand 'fall of the regime' in Beirut". In: *Al Jazeera*. URL: https://www.aljazeera.com/news/2019/10/lebanon-protests-thousands-demand-fall-regime-beirut-191018101342776.html, 11.11.2019.

Algan, Yann and Pierre Cahuc (2010). "Inherited trust and growth". In: *American Economic Review* 100.5, pp. 2060–92.

Bardhan, Pranab (2000). "Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India". In: *Economic Development and Cultural Change* 48.4, pp. 847–865.

Beinin, Joel and Joe Stork (1997). *Political Islam: Essays from middle east report*. Berkeley: University of California Press.

Bjørnskov, Christian (2007). "Determinants of generalized trust: A cross-country comparison". In: *Public Choice* 130.1-2, pp. 1–21.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch (2014). "hroot: Hamburg registration and organization online tool". In: *European Economic Review* 71, pp. 117–120.

Cavalcanti, Carina, Felix Schläpfer, and Bernhard Schmid (2010). "Public Participation and Willingness to Cooperate in Common-pool Resource Management: A Field Experiment with Fishing Communities in Brazil". In: *Ecological Economics* 69.3, pp. 613–622.

Charness, Gary, Uri Gneezy, and Alex Imas (2013). "Experimental methods: Eliciting risk preferences". In: *Journal of Economic Behavior & Organization* 87, pp. 43–51.

Chaudhuri, Ananish (2011). "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the literature". In: *Experimental Economics* 14.1, pp. 47–83.

Chuah, Swee-Hoon, Robert Hoffmann, Martin Jones, and Geoffrey Williams (2007). "Do cultures clash? Evidence from cross-national ultimatum game experiments". In: *Journal of Economic Behavior & Organization* 64.1, pp. 35–48.

— (2009). "An economic anatomy of culture: Attitudes and behaviour in inter-and intra-national ultimatum game experiments". In: *Journal of Economic Psychology* 30.5, pp. 732–744.

Church, Timothy, Marcia Katigbak, Kenneth Locke, Hengsheng Zhang, Jiliang Shen, José de Jesús Vargas-Flores, Joselina Ibáñez-Reyes, Junko Tanaka-Matsumi, Guy

Curtis, and Helena Cabrera (2013). "Need satisfaction and well-being: Testing self-determination theory in eight cultures". In: *Journal of Cross-Cultural Psychology* 44.4, pp. 507–534.

Ciftci, Sabri (2010). "Modernization, Islam, or social capital: what explains attitudes toward democracy in the Muslim world?" In: *Comparative Political Studies* 43.11, pp. 1442–1470.

Ciftci, Sabri, F Michael Wuthrich, and Ammar Shamaileh (2019). "Islam, Religious Outlooks, and Support for Democracy". In: *Political Research Quarterly* 72.2, pp. 435–449.

Cohn, Alain, Michel André Maréchal, David Tannenbaum, and Christian Lukas Zünd (2019). "Civic honesty around the globe". In: *Science* 365.6448, pp. 70–73.

Collins, Kathleen and Erica Owen (2012). "Islamic religiosity and regime preferences: Explaining support for democracy and political Islam in Central Asia and the Caucasus". In: *Political Research Quarterly* 65.3, pp. 499–515.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross (1996). "Cooperation Without Reputation: Experimental Evidence from Prisoner's Dilemma Games". In: *Games and Economic Behavior* 12.2, pp. 187–218.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy". In: *American Economic Review* 100.5, pp. 2205–29.

Dannenberg, Astrid and Carlo Gallier (2020). "The choice of institutions to solve cooperation problems: a survey of experimental research". In: *Experimental Economics* 23, 716—749.

Dannenberg, Astrid, Corina Haita-Falah, and Sonja Zitzelsberger (2020). "Voting on the threat of exclusion in a public goods experiment". In: *Experimental Economics* 23, pp. 84–109.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner (2011). "Individual risk attitudes: Measurement, determinants, and behavioral consequences". In: *Journal of the European Economic Association* 9.3, pp. 522–550.

Economist Intelligence Unit (2019). *Democracy Index 2018: Me too?*

Eickelman, Dale F and James Piscatori (2018). *Muslim Politics*. Vol. 4. Princeton University Press.

Entelis, John Pierre (1997). *Islam, democracy, and the state in North Africa*. Indiana University Press.

Esposito, John L and James P Piscatori (1991). "Democratization and islam". In: *Middle East Journal* 45.3, pp. 427–440.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). "Global evidence on economic preferences". In: *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Feldman, Stanley and Karen Stenner (1997). "Perceived threat and authoritarianism". In: *Political Psychology* 18.4, pp. 741–770.

Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments". In: *Experimental Economics* 10.2, pp. 171–178.

Frey, Bruno S, Matthias Benz, and Alois Stutzer (2004). "Introducing procedural utility: Not only what, but also how matters". In: *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pp. 377–401.

Fukuyama, Francis (1992). *The End of History and the Last Man*. New York: Avon.

— (1995). *Trust: The social virtues and the creation of prosperity*. Vol. 99. New York: Free Press.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni (2005). "Cross-cultural differences in norm enforcement". In: *Behavioral and Brain Sciences* 28.6, pp. 822–823.

Gallier, Carlo (2020). "Democracy and compliance in public goods games". In: *European Economic Review* 121, p. 103346.

Gallier, Carlo, Martin Kesternich, and Bodo Sturm (2017). "Voting for burden sharing rules in public goods games". In: *Environmental and Resource Economics* 67.3, pp. 535–557.

Gallier, Carlo, Jörg Langbein, and Colin Vance (2018). "Non-binding Restrictions, Cooperation, and Coral Reef Protection: Experimental Evidence from Indonesian Fishing Communities". In: *Ecological economics* 150, pp. 62–71.

Grossman, Guy and Delia Baldassarri (2012). "The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda". In: *American Journal of Political Science* 56.4, pp. 964–985.

Güth, Werner and Martin G Kocher (2014). "More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature". In: *Journal of Economic Behavior & Organization* 108, pp. 396–409.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982). "An experimental analysis of ultimatum bargaining". In: *Journal of economic behavior & organization* 3.4, pp. 367–388.

Hamman, John R., Roberto A. Weber, and Jonathan Woon (2011). "An Experimental Investigation of Electoral Delegation and the Provision of Public Goods". In: *American Journal of Political Science* 55.4, pp. 738–752.

Helliwell, John F and Robert D Putnam (2004). "The social context of well–being". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359.1449, pp. 1435–1446.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001). "In search of homo economicus: behavioral experiments in 15 small-scale societies". In: *American Economic Review* 91.2, pp. 73–78.

Henrich, Joseph, Steven J Heine, and Ara Norenzayan (2010). "Most people are not WEIRD". In: *Nature* 466.7302, p. 29.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako,

Natalie Henrich, et al. (2006). "Costly punishment across human societies". In: *Science* 312.5781, pp. 1767–1770.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008). "Antisocial punishment across societies". In: *Science* 319.5868, pp. 1362–1367.

Hofmann, Steven Ryan (2004). "Islam and democracy: Micro-level indications of compatibility". In: *Comparative Political Studies* 37.6, pp. 652–676.

Huntington, Samuel (1996). *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Schuster.

Inglehart, R, C Haerpfer, A Moreno, C Welzel, K Kizilova, J Diez-Medrano, M Lagos, P Norris, E Ponarin, and B Puranen (2014). *World Values Survey: Round Six - Country-Pooled Datafile Version*.

Inglehart, Ronald (2004). *Human beliefs and values: A cross-cultural sourcebook based on the 1999-2002 values surveys*. Siglo XXI.

Kedourie, Elie (1994). *Democracy and Arab political culture*. London: Frank Cass.

Knack, Stephen and Philip Keefer (1997). "Does social capital have an economic payoff? A cross-country investigation". In: *The Quarterly journal of economics* 112.4, pp. 1251–1288.

Kumar, Deepa (2010). "Framing Islam: The resurgence of Orientalism during the Bush II era". In: *Journal of Communication Inquiry* 34.3, pp. 254–277.

Ledyard, John O (1994). "Public goods: A survey of experimental research". In:

Lipset, SEYMOUR MARTIN (1994). "THE SOCIAL REQUISITES OF DEMOCRACY REVISITED". In: *American Sociological Review* 59.1, pp. 1–22.

Litina, Anastasia (2016). "Natural land productivity, cooperation and comparative development". In: *Journal of Economic Growth* 21.4, pp. 351–408.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran (2014). "Self-Organization for Collective Action: An Experimental Study of Voting on Sanction Regimes". In: *Review of Economic Studies* 81.1, pp. 301–324.

Midlarsky, Manus I (1998). "Democracy and Islam: Implications for civilizational conflict and the democratic peace". In: *International studies quarterly* 42.3, pp. 485–511.

Olken, Benjamin A. (2010). "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia". In: *American Political Science Review* 104.02, pp. 243–267.

Reeskens, Tim (2013). "But who are those "most people" that can be trusted? Evaluating the radius of trust across 29 European societies". In: *Social indicators research* 114.2, pp. 703–722.

Robbins, Michael (2015). "After the Arab Spring: People Still Want Democracy". In: *Journal of Democracy* 26.4, pp. 80–89.

Rose, Richard (2000). "How much does social capital add to individual health?" In: *Social science & Medicine* 51.9, pp. 1421–1435.

Roth, Alvin E, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991). "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study". In: *The American economic review*, pp. 1068–1095.

Ryan, Richard M and Edward L Deci (2000). "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." In: *American Psychologist* 55.1, p. 68.

Scharpf, Fritz W (1999). *Governing in Europe: Effective and democratic?* Oxford University Press.

Stenner, Karen (2005). *The authoritarian dynamic*. Cambridge University Press.

Stuetzer, Michael, David B Audretsch, Martin Obschonka, Samuel D Gosling, Peter J Rentfrow, and Jeff Potter (2018). "Entrepreneurship culture, knowledge spillovers and the growth of regions". In: *Regional Studies* 52.5, pp. 608–618.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher (2010). "Choosing the carrot or the stick? Endogenous Institutional Choice In Social Dilemma Situations". In: *The Review of Economic Studies* 77.4, pp. 1540–1566.

Tayekh, R (2001). "Faith-Based Initiatives: Can Islam Bring Democracy to the Middle East?" In: *Foreign Policy, November/December*.

Tessler, Mark (2002). "Islam and Democracy in the Middle East: The Impact of Religious Orientations on Attitudes toward Democracy in Four Arab Countries". In: *Comparative Politics* 34.3, pp. 337–354.

Transparency International (2020). *Corruption Perceptions Index 2019*. Tech. rep. Transparency International, International Secretariat, Berlin.

Tyran, Jean-Robert and Lars P. Feld (2006). "Achieving Compliance When Legal Sanctions Are Non-deterrent". In: *The Scandinavian Journal of Economics* 108.1, pp. 135–156.

Uslaner, Eric M and Mitchell Brown (2005). "Inequality, trust, and civic engagement". In: *American politics research* 33.6, pp. 868–894.

Vereshchagina, Galina and Hugo A Hopenhayn (2009). "Risk taking by entrepreneurs". In: *American Economic Review* 99.5, pp. 1808–30.

Vollan, Björn, Andreas Landmann, Yexin Zhou, Biliang Hu, and Carsten Herrmann-Pillath (2017). "Cooperation and authoritarian values: An experimental study in China". In: *European Economic Review* 93, pp. 90–105.

Wike, Richard, Katie Simmons, Bruce Stokes, and Janell Fetterolf (2017). *Globally, Broad Support for Representative and Direct Democracy*. Tech. rep. Pew Research Center.

Zelmer, Jennifer (2003). "Linear public goods experiments: A meta-analysis". In: *Experimental Economics* 6.3, pp. 299–310.

C Appendix to Chapter 3

C.1 Tables

TABLE 3.10: Cooperation and Religion – Prisoners' Dilemma

	(1) cooperation	(2) modification
Muslim	-0.030	0.120
	(0.060)	(0.105)
Devout	-0.071*	-0.037
	(0.036)	(0.060)
Laicist	0.018	-0.107
	(0.021)	(0.063)
Age	-0.001	-0.012
	(0.011)	(0.010)
Female	0.036	-0.167**
	(0.047)	(0.060)
Round	-0.033***	
	(0.005)	
Constant	0.489*	0.721**
	(0.210)	(0.230)
N	1790	179
R^2	0.051	0.048

Standard errors clustered at session level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3.11: Cooperation and Religion – The Effect of Democracy

	Dependent variable: cooperation in Stage 3			
	(1)	(2)	(3)	(4)
	all	Muslim	Christian	devout
EndoMod	-0.055	-0.078	0.034	-0.047
	(0.094)	(0.093)	(0.101)	(0.211)
ExoNot	-0.539***	-0.529***	-0.648***	-0.531**
	(0.117)	(0.124)	(0.071)	(0.174)
EndoNot	-0.493***	-0.495***	-0.523***	-0.379**
	(0.068)	(0.072)	(0.075)	(0.146)
yes-Voter	0.117***	0.096**	0.256**	0.149
	(0.031)	(0.036)	(0.079)	(0.116)
Muslim	-0.055			0.410**
	(0.035)			(0.169)
Round	-0.010**	-0.009*	-0.018	-0.013
	(0.003)	(0.004)	(0.013)	(0.008)
Constant	0.826***	0.767***	0.916***	0.409*
	(0.078)	(0.085)	(0.227)	(0.209)
N	1800	1580	220	400
R^2	0.281	0.260	0.441	0.242

Standard errors (in parentheses) clustered at session level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3.12: Obedience and Trust – The Effect of Democracy

	Dependent variable: cooperation in Stage 3			
	(1)	(2)	(3)	(4)
	all	obey=1 only	all	trust=1 only
Obey	0.054***			
	(0.013)			
Trust			0.014	
			(0.050)	
EndoMod	-0.052	-0.176*	-0.059	0.208*
	(0.098)	(0.076)	(0.098)	(0.107)
ExoNot	-0.531***	-0.515***	-0.544***	-0.370***
	(0.117)	(0.122)	(0.117)	(0.106)
EndoNot	-0.493***	-0.500***	-0.499***	-0.393**
	(0.070)	(0.048)	(0.070)	(0.158)
Modification	0.111***	0.203***	0.114***	0.232**
	(0.030)	(0.034)	(0.031)	(0.089)
Round	-0.010**	-0.007*	-0.010**	-0.027*
	(0.003)	(0.004)	(0.003)	(0.012)
Constant	0.744***	0.740***	0.781***	0.868***
	(0.080)	(0.075)	(0.079)	(0.153)
N	1800	1140	1800	330
R^2	0.282	0.273	0.279	0.388

Standard errors (in parentheses) clustered at session level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3.13: Overview of Group Level Data – Egypt

	Endogenous Condition		*Exogenous Condition*		
Voteshare	EndoMod	EndoNot	ExoMod	ExoNot	Total
	Number of groups in each outcome:				
0	0	4	2	1	7
1	1	2	7	2	12
2	2	6	3	4	15
3	4	2	0	2	8
4	2	0	1	0	3
Total	9	14	13	9	45
	Cooperation rates in Part 3 (in percent):				
0		5.11	37.5	6.82	
1	68.18	26.14	52.27	5.68	
2	44.32	15.53	84.09	15.34	
3	62.50	23.86		14.77	
4	61.36		90.91		
Average	58.84	15.26	60.31	12.12	

TABLE 3.14: Overview of Group Level Data – Germany

	Endogenous Condition		*Exogenous Condition*		
Voteshare	EndoMod	EndoNot	ExoMod	ExoNot	Total
	Number of groups in each outcome:				
0	0	1	0	0	1
1	0	1	0	1	2
2	4	2	4	5	15
3	6	2	2	2	12
4	2	0	2	1	5
Total	12	6	8	9	35
	Cooperation rates in Part 3 (in percent):				
0		0.0			
1		2.8		100.0	
2	83.8	6.3	40.6	30.5	
3	95.0	8.8	75.0	36.3	
4	9.8		97.5	37.5	
Average	91.7	5.5	63.4	40.3	

C.2 Instructions

Welcome to the experimental lab. In this economic experiment, you will make decisions, which influence how much money you earn for your participation. Please keep in mind that from now on you are not allowed to communicate with anyone other than the lab personnel. If a question arises please show your hand and we will contact you. You must not use a phone, tablet or similar device throughout the entire session. Please note that any act of non-compliance with these rules may lead to your exclusion from all payments. Every decision you will make during the experiment will be treated anonymously and cannot be linked to your identity.

Please read these instructions carefully. The following experiment has two parts and you receive the instructions for the second part after the first is completed. Both parts consist of a game that is played for ten rounds. You earn points in these games; the amount of points you earn depends on your own and on others' choices. At the end, one round from the first and one round from the second part are randomly selected. You are paid your earnings from these two rounds. Points are converted into Egyptian pounds at a rate of 10 points = 10 EGP.

First of all you are now randomly divided into groups of four and every group member receives a player ID between 1 and 4. Both the group composition as well as all player IDs remain unchanged throughout the entire experiment.
Example: You are player 2 and form a group with the players 1, 3, and 4.

Part 1

In this part you play ten rounds of a game (**Game 1**) together with one of your other three group members. This other player is randomly chosen in every round and you will be notified at the end of the round who your partner was. **In this game you can decide between the options A and B in each round.** Your partner simultaneously chooses one of the options. While you make your decision, you do not know what your partner chooses. Your income in each round of game 1 is calculated in the following way:

If both you and your partner choose option A you both earn 50 points.
If you choose option A and your partner chooses B, then you earn 30 points and your partner earns 60.
If you choose option B and your partner chooses A, then you earn 60 points and your partner earns 30.
If both you and your partner choose option B you both earn 40 points.

After each round you will see the chosen option of your partner and of the other group members on your computer screen. Table 1 gives an overview of your earnings per round in game 1.

Your Choice	Your Partner's Choice	
	A	B
A	50	30
B	60	40

TABLE 3.15: Game 1

Part 2

Part 2 of the experiment starts with a vote. **Every group elects one of their members as their speaker in a secret ballot. This speaker can decide which game your group will play for ten more rounds.** The choice is between Game 1 (as known from Part 1) and Game 2. In Game 2 you can again choose between options A and B and your income is calculated in the following way:

If both you and your partner choose option A you both earn 50 points.
If you choose option A and your partner chooses B, then you earn 30 points and your partner earns 45.
If you choose option B and your partner chooses A, then you earn 45 points and your partner earns 30.
If both you and your partner choose option B you both earn 40 points.

Your Choice	Your Partner's Choice	
	A	B
A	50	30
B	45	40

TABLE 3.16: Game 2

At first you must now indicate which game you would choose for your group in case you become speaker. This decision is secret until the election of the speaker is completed. For this you vote for one other group member. You cannot vote for yourself. In case of a tie one of the players with the highest amount of votes is randomly chosen as speaker. The speaker's choice of game becomes binding for the

entire group. **However, this choice is only implemented with a probability of 50 percent.** If the speaker's game is not implemented the computer randomly selects Game 1 or 2. **Both games are equally likely to be chosen in this case.**

You will be informed about who was elected as speaker, which game he/she preferred, if this choice was considered and if not which game your group will play in part 2 and **you play this game for ten rounds.** Again, you are informed about your partner's and other group members' choices after each round.

Subsequently we are going to ask you to fill out a short questionnaire, which has no influence on your income, and determine the two rounds relevant for the payout.

C.3 Democracy Premium

Calculation of policy effects in round 11, following the identification strategy by DFP (2010).

Egypt

Total policy effect: [54.5(11/36) + 64.0(25/36)] - [30.6(36/56) + 40.0(20/56)] = 27.14.

Selection effect: 30.6(11/36 - 13/24) + 40.0(25/36 - 20/56) = 3.17.

Exogenous treatment effect: (11/36)(57.1 - 25.0) + (25/36)(88.2 - 12.5) = 62.38

Democracy premium: 23.97 - 62.38 = -38.41

Germany

Total policy effect: [64.3(14/48) + 94.1(34/48)] - [23.1(13/24) + 27.3(11/24)] = 60.38.

Selection effect: 23.1(14/48 - 13/24) + 27.3(34/48 - 11/24)] = 1.05.

Exogenous treatment effect: (14/48)(40.0 - 46.7) + (34/48)(81.8 - 61.9) = 12.14.

Democracy premium: 59.33 - 12.14 = 47.19.

C.4 Additional Questionnaire in Egypt

C.4.1 Questionnaire Items

- How old are you?
- What is your gender?
- What is your monthly income?
- Were you born in this country or are you an immigrant ?
- Do you belong to a religion or religious denomination? If yes, which one?
- How many times per day do you pray on average?
- Do you observe the religious fasting regulations?
- With which of the following statements do you agree?

- I have a high willingness to take risks.
- Generally speaking, most people can be trusted.
- Religion and state governance should be separate from each other.
- Schools should teach children to obey authority.
- Our leaders know what is best for us.
- Most people can learn to be leaders – it's not a matter of birth.
- I am very persevering — and I usually accomplish what I set out to do.
- I feel good when I cooperate with others.
- Young people today do not have enough respect for traditional values.
- People don't know the difference between right and wrong anymore.
- I respect the majority's wishes in groups of which I am a member.

C.4.2 Questionnaire Results Not Discussed in Main Analysis

Table 3.17 and Figure 3.7 give an overview of the questionnaire results from Egypt.

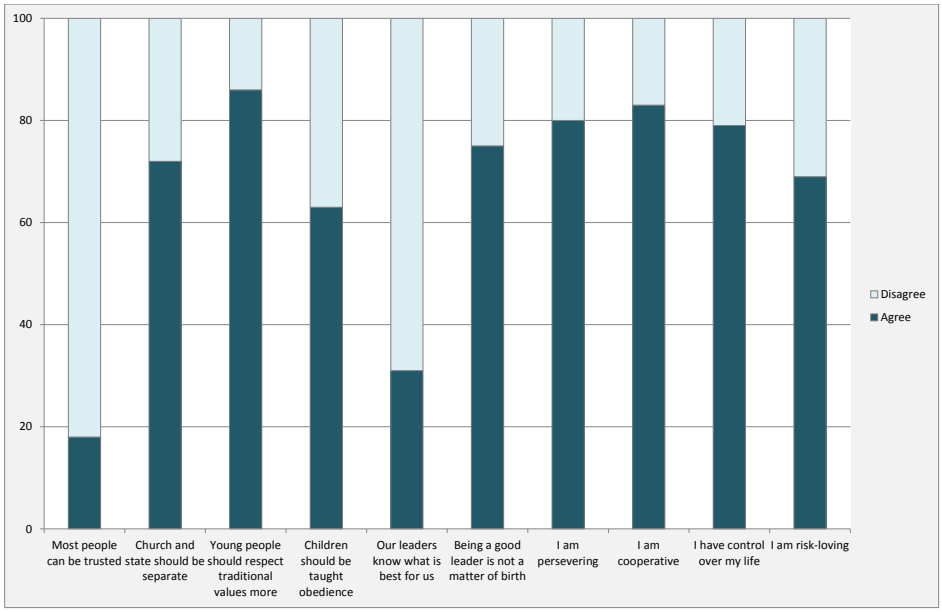
TABLE 3.17: Questionnaire Answers Egypt

	Share agreed (%)			
	All	Muslim	Christian	*p*-value
Separate religion and state	72	70	91	0.04**
Most people can be trusted	18	18	23	0.57
Respect more traditional values	86	85	86	0.91
Schools should teach obedience	63	62	73	0.33
Our leaders know what is best for us	31	31	27	0.68
Most people can learn to be leaders	75	75	77	0.79
Perseverance	80	79	86	0.43
Desire to cooperate	83	82	86	0.68
Locus of control	79	80	77	0.79
I am willing to take risks	69	68	77	0.40

Note: Sample size is n = 180. *p*-values are obtained from Mann-Whitney tests for differences in the means between Muslim and Christian subjects.

Religiosity 88 percent of the sample identified as Muslims; almost half of those (44 percent) said they prayed five times a day or more. The majority of Christians prayed once or twice a day and mostly adhered to their fasting requirements (as opposed to not fasting or strictly fasting). 63 percent of the Muslims followed fasting regulations strictly; and 24 percent of Muslims and 5 percent of Christians were both strict in fasting *and* praying, those are henceforth characterized as *devout*. 70 percent of Muslims and 91 percent of Christians agreed with the general concept of separating religion and state ($p < 0.04$).

FIGURE 3.7: Questionnaire Answers



Risk and Self-determination 69 percent of the subjects self-assessed themselves as having a high willingness to take risks. This self-reported measure of risk-aversion was found to be a good predictor of real-life risk preferences (Dohmen et al., 2011; Charness, Gneezy, and Imas, 2013). Previous studies hypothesized lower individual risk-aversion is associated with higher levels of entrepreneurship, which can be beneficial for economic growth (Vereshchagina and Hopenhayn, 2009; Stuetzer et al., 2018). The relatively high share of subjects reporting to be risk-loving, or at least not at all risk-averse, is consistent with findings in other studies reporting that Middle Eastern and North African populations have the highest risk tolerance in the world (Falk et al., 2018, p.1664). Risk aversion has no explanatory power in the experiment. A high share of subjects (79 percent) saw themselves as being in control over their own life and as persevering (80 percent).

Authoritarianism and Leadership 63 percent agreed that it is important to teach children to obey authority. Child-rearing values are supposedly the strongest predictor of overall authoritarian predispositions (Feldman and Stenner, 1997; Stenner, 2005). I therefore use *obedience* as the explanatory variable to investigate hypothesis 5. A large majority (86 percent) also agreed that "young people today do not respect traditional values enough". Put together, the participants display a rather conservative mindset, especially when considering that the pool consists of young social science students in an urban environment.

On the other hand, 75 percent of subjects agreed with the meritocratic approach that being a good leader is not a matter of birth. Less than one third of the participants agreed that "our leaders know what is best for us", which is not surprising given the latent dissatisfaction of the Egyptian people with their government.

Trust and Cooperation Only 18 percent of the experiment population agreed to the statement that "generally speaking, most people can be trusted". The question is part of the World Values Survey by Inglehart et al. (2014), who found a comparable result in a large scale survey (1,523 participants) in Egypt in 2012, where 21.5 percent agreed to the statement. The level of trust in Egypt significantly decreased within one decade: an earlier wave of the World Values Survey reported a share of 38 percent agreeing to the statement (Inglehart, 2004). For comparison: in 2013, the average level of agreement in the EU was 35 percent (Reeskens, 2013). Trust does not constitute a preference, but rather a belief (Falk et al., 2018). A large strand of literature suggests that high levels of trust are positively connected to a host of development objectives such as economic growth, productivity, happiness, health, and equality (Fukuyama, 1995; Knack and Keefer, 1997; Rose, 2000; Helliwell and Putnam, 2004; Uslaner and Brown, 2005; Algan and Cahuc, 2010; Litina, 2016). Regarding potential determinants of trust, more hierarchical religions like Islam (or Catholicism) are associated with lower trust levels, ceteris paribus (Bjørnskov, 2007).

83 percent of all subjects expressed a desire to cooperate with others. However, there is no significant relationship between stating this preference and acting on it. The questionnaire answer is insignificant as a predictor of cooperation in the prisoners' dilemma and modification preferences.

Chapter 4

In for a Penny, in for a Pound: Accession Costs and Voting Rules in International Organizations

Authors Fanny Schories, Stephan Michel

Abstract International organizations face the problem of balancing the benefits of enlargement with the risk of becoming less effective in club good provision whenever there is uncertainty about the productivity of candidate states. We use a screening model with accession costs and voting rules as choice variables to explain how this balance should be found. In contrast to previous literature, we explicitly consider the interplay between both institutional features. Accession costs are used to extract rents for the incumbents whenever possible, whereas the voting rule is the efficient screening device also when accession costs are wasteful spending. Simple majority is the optimal voting rule when candidates are expected to be very productive or very unproductive. Unanimity is optimally used only for medium productivity.

Keywords Asymmetric Information, Club Good, European Central Bank, International Organization, Majority Rule, Unanimity Rule.

JEL Classification D02, D71, D82, F53, F55, K33.

4.1 Introduction

> *When we talk about a large Europe, then immediately the question as to the EU's ability to act arises. Obviously, a large Europe mustn't put the brakes on the ambitions of others who want to work together even more closely. Instead of categorising people as good or bad Europeans, we should take note of the fact that the objective of ever closer union for the countries of Europe is not shared to the same extent by all Member States.*
>
> Heiko Maas, German Minister of Foreign Affairs, 13.06.2018.

International organizations (IOs) make use of a wide variety of voting rules in their main decision-making bodies: Some have the strong requirement of unanimity, in others choices can be made via simple or qualified majorities. It is not obvious what causes the variation in voting rules across and within IOs. The founding members of an IO have to settle on an institutional design to govern their future operations: Both a procedure to decide which actions should be taken by the organization as well as criteria for the accession of new members have to be determined. Naturally, the effect of a given voting rule depends on who votes. Accession terms influence just this. We therefore argue in this paper that the two features should be analyzed jointly. The screening model developed in the paper shows which rule combinations are preferable for the founding members given their expectations of future IO enlargement.

For example, as a heterogeneous club, the Eurozone provides a common good to its members: a single currency. Countries substituting their national currency with the Euro effectively give up the possibility of conducting independent monetary policy. Instead they rely on the European Central Bank (ECB), whose primary objective is price stability defined as an inflation target of less than but close to 2% (Article 127, Treaty on the Functioning of the European Union (TFEU)). Because the member states' economies can be hit by asymmetric shocks – e.g. differing between export- and import-oriented countries – preferences about monetary policy may vary over time. The original eleven Euro countries were sufficiently similar to rarely disagree on appropriate policy measures such that the voting rule did not matter in the initially small club. Indeed, in the first years, the Governing Council decided by consensus (Riboni and Ruge-Murcia, 2010). However, Eurozone enlargement was politically desired from the onset, such that the institutions were designed in anticipation of a more heterogeneous union. A non-unanimous decision was officially announced for the first time in September 2012, after the Eastern enlargement of the Eurozone had taken place and the debt crisis was hitting countries in dissimilar ways. The decision-making rule used in the Governing Board today is a rotating majoritarian system, thus not every member's preferences will be considered in a given period. We argue that this is directly related to the strictness of the accession criterion. In the case of the Eurozone, the Maastricht criteria determine whether a country can adopt the Euro. The ECB reportedly took a "relaxed position" (Jonas,

2006, p.333) about these criteria in the past. And, as was made public during the Euro crisis starting in 2010, Greece – and in fact several other Eurzone economies – never met the criteria as they are written in the Treaty. Nevertheless, Greece was allowed to become the first candidate to join the Euro after its inception. Neither were the criteria of the stability and growth pact rigorously enforced once a state had become an EMU member (Irlenbusch and Sutter, 2006).

The literature so far considered accession criteria and voting rules separately and argued that IOs ask for concessions from prospective new members in the form of domestic policy adjustments (Koremenos, Lipson, and Snidal, 2001). Accession terms as a costly signal can resolve uncertainty about states' types and lead to a separating equilibrium, where "good" types join and "bad" ones do not (Kydd, 2001). Our model adopts this argument but relates the strategic design of accession costs to the policy-making procedure used within the IO. The concern that new member states reverse policy decisions in their favor if given too much voting power is not new (Shackleton and Laffan, 1996), but we are the first to explicitly model this trade-off.

We address the gap in the literature with the help of an economic model of the trade-off between balancing the benefits of enlargement with the risk of diluting the club good in the presence of uncertainty about the preferences of candidate members. We are proposing an analytical framework to assess what combination of accession conditions and voting rules is optimal for the incumbents of an IO in a given situation. We build a game-theoretic screening model of club good provision under uncertainty. Uncertainty about the potential entrant comes in the form of stochastic variation in the benefits of the good (high or low) and in the candidate being of a type (high or low productivity) that is not publicly known. The founding IO members simultaneously determine the amount a new member has to pay to join their club and whether to aggregate members' preferences about the club good provision using unanimity or a simple majority rule. They face a fundamental trade-off: A stricter voting rule makes it more likely that only Pareto-efficient decisions are taken but at the cost of sometimes foregoing actions which would increase total welfare. Low accession costs make candidates more likely to join and a larger membership increases the benefits for every member. However, low accession costs pose the risk that also unproductive candidates are admitted.

Our main result is that the voting rule is the more efficient screening device: a simple majority rule induces candidate states to self-select into the IO only when they are of the good types, making costly entry barriers obsolete as a screening device. However, when we consider accession costs not as wasteful spending but as a direct transfer to the incumbents, they can be used for rent-extraction and make the admission of candidates, *ceteris paribus*, more attractive. Thus, accession costs are solely a tool for rent-extraction and not a screening device.

The following section 2 reviews the existing literature. Section 3 shows our main contribution in setting out the baseline model with wasteful accession costs as well

as a model extension where accession costs are modeled as direct transfers to the incumbents. A final section discusses the results and puts them in a more general perspective.

4.2 Related Literature

We follow Watson (2004), Ahrens, Hoen, and Ohr (2005), and Ahrens, Ohr, and Zeddies (2007), who analyze the EU as a club. The goods provided by this club are of little rivalry in consumption and exclusively available to member states. The motivating club good example in this paper is the common currency and price stability produced through the Economic and Monetary Union (EMU), whose members are a subset of the EU. Watson (2004) considers the EMU as a club in which congestion poses a natural limit to the optimal number of members. This feature is not present in our model, as we focus on the extensive margin of membership. Instead, we follow suggestions by Mundell (1973a, 1973b) and assume positive returns to scale.

A number of studies have shown that voting behavior in the ECB board reflects national preferences, which will be a central assumption in the model (Cancelo, Varela, and Sánchez-Santos, 2011; Hayo and Méon, 2013; Moschella and Diodati, 2019). Since individual preferences about monetary policy can vary, the decision-making committee has to use a rule with which these preferences are aggregated. Our focus is on the comparison of the two most basic (and extreme) voting rules: unanimity and simple majority. Riboni and Ruge-Murcia (2010) develop a dynamic voting game to study decision-making in a heterogeneous central bank committee. They show that consensus (or unanimity rule) fits most of the actual observed policy decisions by central banks. Unanimity ensures a Pareto-superior outcome and thus faces no enforcement problem but entails high decision-making cost since every member state can veto a proposal (Kirchner, 2012; Posner and Sykes, 2014). Majority rule mitigates the hold-out problem and ensures greater responsiveness of the IO, but potentially allows exploitation of the minority. Blake and Payton (2015) thus claim that unanimity generally makes membership in an IO more attractive, a finding which is supported by the results of this paper. Dougherty and Edward (2012) and Dougherty et al. (2014) show that majority rule may be more successful than unanimity in attaining Pareto-optimal outcomes if the policy space is multidimensional. Our paper shows that a comparable conclusion can be reached even for one-dimensional decisions. Empirically, Blake and Payton (2015) and Hooghe and Marks (2015) find that IOs with more members tend to have smaller majority requirements. Our model will provide one potential explanation why this is the case: increased preference heterogeneity makes consensus less likely with increasing membership.

Kandogan (2000) argues that in a homogeneous club the voting rule is of minor importance; unanimity functions well. However, with every enlargement wave the organization becomes more heterogeneous, such that the voting rule is the crucial aspect in institutional design to adapt according to the accession of poorer states.

Specifically, it is suggested that the majority requirement should be lowered before enlargement. Our paper comes to a similar conclusion, but emphasizes also the flip side of the argument: not only is the optimal voting rule influenced by the composition of the club, but it also influences who joins it in the first place. Gray, Lindstädt, and Slapin (2017) study how varying degrees of preference heterogeneity before and after enlargement influence the dynamics of IO decision-making in an agent-based model. Similar to the present paper, the candidate state is privately informed of its own type, which can be good or bad. Gray, Lindstädt, and Slapin (2017) are grounded in the literature on signaling models, in which the informed party – the candidate – makes an offer regarding the IO accession terms. Conversely, the present paper applies a screening model, in which the uninformed part – the IO – offers the candidate a combination of voting rule and accession costs. We believe this to be a closer representation of the negotiation procedures and allocation of bargaining power in the real world.

In general, an IO like any other club is willing to include only those new members that bring efficiency gains to the incumbents. This may be hard to predict ex ante, as in our model setting with uncertainty about the productivity of other states. Koremenos, Lipson, and Snidal (2001) argue that candidates potentially want to misrepresent their true type, which can be mitigated by purposefully designed accession terms. As a consequence membership becomes less open the higher the uncertainty about others' preferences. A recurring point in the literature on the (optimal) size of nations and IOs is the broader-deeper trade-off, suggesting that due to gridlock and collective action problems an organization cannot have a large membership and deep cooperation at the same time (Downs, Rocke, and Barsoom, 1998; Alesina and Spolaore, 1997; Alesina and Spolaore, 2003; Stone, Slantchev, and London, 2008). Gilligan (2004) shows that the trade-off disappears once the assumption that all members have to set identical policies is relaxed. While individual policy levels may be plausible for some IOs, diverging policies are not an option for the large set of IOs whose purpose is to set standards or prices. The former are a way of solving international coordination problems, in which the members' benefit arises exactly from the fact that actions are harmonized. The latter consist of interest rates for a common currency or (the absence of) tariffs in a free trade area. By definition, a monetary union like the Eurozone cannot have individual monetary policy conducted by its member states. As such, we follow Maggi and Morelli (2006) who assume that the collective action of the IO is only effective if all members participate. To summarize, we proceed from existing studies in two ways: on the one hand by combining the strategic considerations between an IO's choice of voting rule and accession terms; on the other hand by endogenizing the size of the IO in a model of club good provision.

4.3 The Model

4.3.1 Basic Framework

Goods provided by an IO like the Eurozone, e.g., a single currency, can feasibly be considered as club goods that are non-rival in consumption but exclusively available to member states. Specifically, the good provided to members of the "Euro Club" is a stable inflation rate, ensured via the adaptation of the interest rate applicable to the main refinancing operations of the ECB. In a nutshell, raising the interest rate corresponds to a contractionary monetary policy and thus lower inflation. This policy measure is represented by the club good provided in the following model, a policy measure over which the participating member states may have diverging preferences. We suggest a screening model for the provision of the club good with incomplete information and endogenous club membership. In this context we consider different voting rules to aggregate members' preferences. The interaction is a one-shot game between three players = $\{P_1, P_2, P_3\}$. The first two (*incumbents*) form an IO[1] and the third is a potential new member state (*candidate*). The trade-off is that the two incumbents always want to produce the club good, but the candidate might not. With majority rule, they could outvote a disagreeing third party, but then the candidate could not find it attractive to join. This problem does not exist with unanimity rule, but then the new member might veto many decisions. Therefore, costly accession screens out the unproductive candidates.

To make the argument more formal, we develop the following model with four stages. Initially, the incumbents set up the IO and choose a voting rule. The voting rule of the IO of size $m = 3$ is characterized by a number $r^* \in \{1, 2, 3\}$. Club good provision takes place if and only if at least r^* players vote in favor. To study the effect of two prominent voting rules, we limit attention to simple majority rule, i.e. $r^* = \frac{m+1}{2}$ and unanimity, i.e. $r^* = m$.

In a second stage, the accession costs c are determined. c is the amount that P_3 has to pay in order to become a member. For now, let us assume that c is a sunk cost and of no benefit to the incumbent members. The incumbents do not have to pay for their membership. Upon learning about c and r^*, P_3 decides whether to become a member (pay c) or not. For the final stage an individual benefit parameter is drawn and all IO members vote on whether or not to produce a club good. Since enforcement is perfect, the good is always produced in accordance with the vote outcome. Staying with the example of the common currency, accession to the Eurozone is a full commitment. It is formally regulated by Article 140(3) TFEU that the adoption of the Euro as currency cannot be reversed. And indeed, despite swaying public opinions in several member states, as of yet no country has ever withdrawn from the Euro.

[1] Please note that our model deals with a simplified version of an IO. We are fully aware that two members would not be sufficient to form an IO by the prevailing view in international law. The results would also apply with a generalized version of the model with n incumbents.

We will therefore assume in our theory that potential club members are bound by their accession decision making the Euro a true club good without congestion.

4.3.2 Payoffs

Payoffs depend on players' (high or low) type θ_i^j with $j \in \{L, H\}$, which in turn determines the benefit from club good production, given by parameter $\lambda_j \in \{\underline{\lambda}, \overline{\lambda}\}$, with $\overline{\lambda} > \underline{\lambda}$. Specifically, $\theta_i^j = Prob(\overline{\lambda})$ and $(1 - \theta_i^j) = Prob(\underline{\lambda})$. We assume that θ^H is always equal to 1 and that it is commonly known that both incumbents are of the high type, i.e. $\theta_{1,2} = \theta^H = 1$. It follows by assumption that the high types receive a high benefit with certainty. The candidate could be a high or a low type with equal probability. If P_3 is a high type, then she too will draw $\overline{\lambda}$ with certainty. For low types $0 < \theta^L < 1$ applies and both realizations of λ are possible. We remain agnostic as to the exact origin of this preference divergence – feasible causes are domestic policy concerns such as business cycles or elections; global shocks, to which some economies are more vulnerable than others; or countries' respective trade balances (Matsen and Røisland, 2005). Payoffs are strictly increasing in the size of the IO as represented by the common parameter $a'(m) > 0$. These economies of scale can stem, inter alia, from network effects or shared risks. The outside option to production and membership is worth 1. Individual payoffs of the incumbents and the candidate are then:

$$u_{1,2} = \begin{cases} a(m) \cdot \overline{\lambda} & \text{if club good produced} \\ 1 & \text{otherwise} \end{cases}$$

$$u_3 = \begin{cases} a(m) \cdot \lambda_j - c_3 & \text{if member and club good produced} \\ 1 - c_3 & \text{if member and club good not produced} \\ 1 & \text{otherwise} \end{cases}$$

The analysis focuses on the interesting case in which $\overline{\lambda}$ makes production individually profitable for any $a(m)$ while $\underline{\lambda}$ yields a net loss for player 3. We therefore assume $a(m) \cdot \underline{\lambda} < 1 < a(m) \cdot \overline{\lambda} \quad \forall m$.[2]

To sum up, the incumbents are high types, receive a high benefit from club good production and – due to economies of scale – prefer a larger IO. The candidate can be either a high type as well, or a low type, for whom high or low benefits could occur. These types and their independent benefit draws are a stylized way of modelling countries' inflation preferences as discussed in the previous section: Some high type countries always strictly prefer a monetary policy of high price stability. For the other type the preferences depend on the state of the world in any given period. A less developed economy could for instance prefer higher inflation rates in times of higher output growth. The timing of events can be seen from figure 4.1. Proceeding

[2]Trivial solutions arise if the assumption is not met. For $\overline{\lambda} > \underline{\lambda} > \frac{1}{a(m)}$ every type always favors production and the voting rule is irrelevant. If $\frac{1}{a(m)} > \overline{\lambda} > \underline{\lambda}$ the organization is pointless.

by backwards induction the equilibrium predictions that follow from this model setup are discussed in the following.

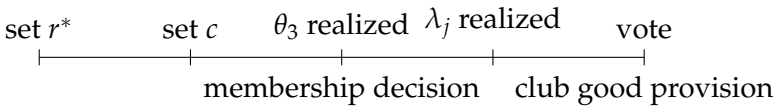


FIGURE 4.1: Model Timing

4.3.3 Equilibrium Strategies

Voting Decision

In the final stage, low types learn their private benefit draw λ_j and all IO members cast their vote on whether to produce the club good. The incumbents and a new high type member vote yes with certainty. A low type member votes yes if they received a high benefit draw ($\overline{\lambda}$) and no if they received a low benefit draw ($\underline{\lambda}$). Any behavior other than truthful voting is always at least weakly dominated. Voting behavior is independent of the voting rule. Because the incumbents are in the majority, the outcome of a simple majority vote is to produce, regardless of the new member's vote. With unanimity, production takes place unless a low type new member receives a low benefit and therefore vetoes.

Accession Decision

When deciding whether or not to join the organization, the entrant knows its own type θ_3^j, the probability distribution of the private benefit draw λ_j, the voting rule and the accession costs of the IO. Under unanimity, any unfavorable decision can be vetoed, whereas under majority rule the incumbents have a majority of votes.

The decision to join is based on a simple cost-benefit calculus: the value of being a member net of accession costs should be larger than the outside option. Thus, a potential entrant will join if the following condition holds

$$E[u_3(\text{join}|r^*, c, \theta_3^j)] > 1 \tag{4.1}$$

Candidates base their decision on the accession costs versus expected benefit from membership, which in turn depends on the voting rule. Since the new entrant knows her own type, we can separately look at the decision for each type at this stage. Let c^H be the maximum amount a high type candidate would pay for accession given any of the two voting rules. Because high types always agree with the incumbents that production is preferable, they do not care about the voting rule and join whenever the club good's benefit outweighs the accession costs:

$$c^H \equiv c \leq a(3)\overline{\lambda} - 1 \tag{4.2}$$

For a low type the voting rule does matter. Let c^{LU} be the maximum amount a low type candidate would pay for accession given unanimity, and c^{LM} the maximum willingness to pay with majority rule. With unanimity, the low type can veto any detrimental production decision. They will never be forced to produce the good if they do not want to, but the expected benefit from membership and hence the willingness to pay is lower than that of a high type, such that:

$$c^{LU} \equiv c \leq \theta^L(a(3)\overline{\lambda} - 1) \tag{4.3}$$

With majority rule, a low type knows it will be outvoted and production always takes place, even in the – for her – unfavorable case that occurs with probability $(1 - \theta^L)$. The maximum willingness to pay becomes

$$c^{LM} \equiv c \leq \theta^L[a(3)(\overline{\lambda} - \underline{\lambda})] + a(3)\underline{\lambda} - 1 \tag{4.4}$$

Note that even for $c = 0$ equation (4.4) may not be satisfied if θ^L and $\underline{\lambda}$ are sufficiently low. A low type candidate can be in a position where she will rationally stay out of an IO with majority rule, regardless of the accession costs. If we denote the lowest θ^L that given $c = 0$, majority rule and $a(3)$, $\overline{\lambda}$, and $\underline{\lambda}$ makes a low type want to join as θ^*, the low type's participation constraint follows directly from (4.4):

$$\theta^* \equiv \theta^L \geq \frac{\frac{1}{a(3)} - \underline{\lambda}}{\overline{\lambda} - \underline{\lambda}} \tag{4.5}$$

This lower bound on θ^L decreases in $a(3)$ and $\overline{\lambda}$.

Accession Costs

In this stage, the incumbents set the accession cost for a potential entrant. They have previously determined the voting rule and are aware that it is equally likely for the candidate to be a high type like themselves, or a potentially unproductive low type. The high type always wants to contribute, but whether the low type will actually wish to contribute becomes known only after the membership decision is made.

Given a simple majority rule is in place, the incumbents can outweigh a negative vote from the entrant and produce the club good no matter what. Since $a(3) > a(2)$, any additional member strictly increases the expected utility of the incumbents and they set $c = 0$. Since we assume that accession costs are wasteful spending, any positive accession costs ($c > 0$) will only reduce the payoffs of potential entrants without any benefit for the incumbents. Thus, setting the accession costs to zero is Pareto-superior.

Under unanimity, only the high type candidates are beneficial to the incumbents with certainty. If the incumbents expect a positive net gain also from a low type candidate, even at the chance it might veto, they again set $c = 0$. If they expect a low

type entrant to make them worse off, they set $c^* = c^{LM} > 0$ such that a high type is still willing to join but a low type is screened out.[3]

Lemma 1 *With unanimity, in case a separating equilibrium where low types stay out and high types join is preferred by the incumbents, they will set the accession cost to* $c^* = c^{LU}$. *Otherwise,* $c = 0$ *is Pareto-superior to any* $c > 0$.

Proof 1 In the appendix.

Voting Rule

The previous section analyzed optimal accession costs and participation decisions taking the voting rule as given. We now go one step further and ask which of the equilibrium rule combinations the incumbents choose at the constitutional stage.[4] The optimal voting rule depends on players' (expected) cost-benefit parameters. We know that in the absence of prohibitive accession costs high types always participate in the IO.

Under simple majority rule, any new member makes the incumbents strictly better off and they never want to screen through accession costs. Low type candidates wish to participate in a majoritarian IO whenever $\theta^L \geq \theta^*$. In this case, majority rule without accession costs is the optimal equilibrium and all candidates join. If $\theta^L < \theta^*$, such that a low type would not join given majority rule, the incumbents are open to accession if they still benefit from its membership even when granting it a veto right. This is the case whenever

$$\theta^{**} \equiv \theta^L \geq \frac{a(2)\overline{\lambda} - 1}{a(3)\overline{\lambda} - 1} \tag{4.6}$$

Then they optimally implement unanimity without accession costs. If $\theta^L < \theta^{**}$, the potential benefits from a low type member are so low that the incumbents are not willing to grant a veto right. Instead, they implement majority rule and zero accession costs: candidates self-select and only high types join. We find that accession costs as a screening device are never used in equilibrium and the optimal voting rule is discontinuous. Majority rule is optimal for very low and very high ranges of θ^L. Unanimity is optimal in the interval $\theta^L \in [\theta^{**}, \theta^*]$. Ultimately, the choice of rules depends on the incumbents' expectations about the productivity of the low type and their prior about the likelihood that the candidate is indeed a low type. The results of this section are summarized in the following proposition. Figure 4.2 illustrates the three possible equilibria.

[3] Note that we assume the candidate to resolve indifference in the direction of not joining. This could be the case if agents were even slightly risk-averse. The qualitative results of the model do not require this assumption.

[4] Since both incumbents are high types we focus on symmetric behavior. Thus, it does not matter with which voting rule they choose the initial voting rule and the accession costs. Problems of infinite regress do not arise in this setting.

Proposition 1 *The incumbents will choose the following equilibrium combinations of accession cost and voting rule.*

$$c = 0 \text{ and } r^* = 3 \text{ if } \theta^L \in [\theta^{**}, \theta^*]$$
$$c = 0 \text{ and } r^* = 2 \text{ otherwise.}$$

Proof 2 In the appendix.

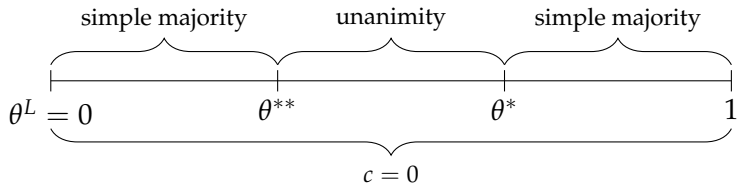


FIGURE 4.2: Equilibrium Rule Choice – Proposition 1

4.3.4 Model Extension: Redistributive Accession Cost

So far, the model assumed accession cost to be wasteful spending and purely a potential screening device. We now change this assumption such that c is redistributed equally among the incumbents. While the voting decision and participation constraints are not affected by this change the equilibrium choice of accession cost and voting rule are.

Accession Cost

From section 4.3.3, we can recall the participation constraints of the different types:

$$c^H \equiv c \leq a(3)\overline{\lambda} - 1$$

$$c^{LU} \equiv c \leq \theta^L(a(3)\overline{\lambda} - 1)$$

$$c^{LM} \equiv c \leq \theta^L[a(3)(\overline{\lambda} - \underline{\lambda})] + a(3)\underline{\lambda} - 1$$

With simple majority, the optimal accession cost is no longer zero. For the incumbents, the optimal accession cost is now either the maximum that a low type candidate would be willing to pay or the maximum that a high type is willing to pay. Setting c somewhere in between can never be optimal, because it leads to lower profits for the incumbents while leaving the membership decision of the candidates unchanged. The trade-off here is the additional utility from the higher accession cost payment of the high type with the foregone benefit of having a low type candidate join. This foregone benefit is composed of the benefit from an additional member and the redistributive accession cost which a low type would have to pay.

For unanimity, the same trade-off applies. However, accepting low types also leads to the risk of non-production in case of a low draw. Thus, the calculus for the incumbents when deciding on the accession cost is the following:

Lemma 2 *The accession cost c^H are optimally chosen by the incumbents if either*

$$\theta^{SM} \equiv \theta^L < \frac{a(2) \cdot \overline{\lambda} - a(3) \cdot (\underline{\lambda} + \frac{1}{2} \cdot \overline{\lambda}) + \frac{1}{2}}{a(3) \cdot (\overline{\lambda} - \underline{\lambda})} \text{ and } r^* = 2$$

or

$$\theta^{UN} \equiv \theta^L < \frac{a(2)\overline{\lambda} - 1}{2 \cdot c^H} + \frac{1}{4} \text{ and } r^* = 3$$

is satisfied.

Proof 3 In the appendix.

Choice of Rules

For the choice of voting rules, we can use the results of the previous section as a starting point. Using Lemma 2 and the fact that whenever c^H is chosen, the outcome for the incumbents is the same independent of the voting rule, we can directly show the following result for the voting rule:[5]

Proposition 2 *With accession costs as transfers to the incumbents they optimally choose the following equilibrium combinations of accession cost and voting rule.*

$$\begin{aligned}
&c = c^{LM} \text{ and } r^* = 2 \text{ if } a(3)(\overline{\lambda} + \underline{\lambda}) > 2 \ \ \& \ \theta^L > \theta^{SM}, \\
&c = c^{LU} \text{ and } r^* = 3 \text{ if } a(3)(\overline{\lambda} + \underline{\lambda}) < 2 \ \ \& \ \theta^L > \theta^{UN}, \\
&c = c^H \text{ and } r^* = 2 \text{ in all other cases.}
\end{aligned}$$

Proof 4 In the appendix.

The results of the model variation are illustrated in figures 4.3 and 4.4. We see that for both voting rules it is only optimal to admit low type candidates if both they individually as well as the IO as a whole are sufficiently productive. The term $a(3)(\overline{\lambda} + \underline{\lambda})$ can be interpreted as how productive a three-member IO would be. Whenever this value increases, membership becomes more attractive for a candidate and thus above the threshold value it is not necessary any more to motivate them with a veto right to make them willing to join.

With regard to the threshold values for θ^L, it becomes clear that candidates that are in expectation more productive are more likely to lead to a three-member IO. In the case of unanimity, the rationale behind the threshold value of θ^L is the underlying veto threat of a low-type candidate. For simple majority, a low θ^L would reduce the

[5]Note that simple majority ($r^* = 2$) and unanimity ($r^* = 3$) are equivalent in the last case.

maximum accession cost a low type is willing to pay to such a low level that setting a higher cost and keeping low types out is more attractive. Note that since θ^{SM} is always bigger than θ^*, the low type would always be willing to join when simple majority with accession costs of c^{LM} is chosen.

To summarize, we find that with accession costs as transfers to the incumbents these do not exclusively rely on the voting rule to screen the candidates, as was the case in the baseline model. Now, they rather extract the maximum amount possible from each candidate.

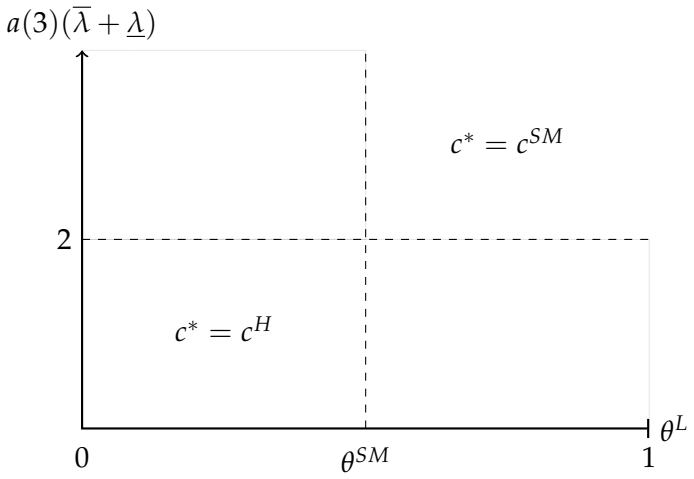


FIGURE 4.3: Optimal Accession Costs - Simple Majority

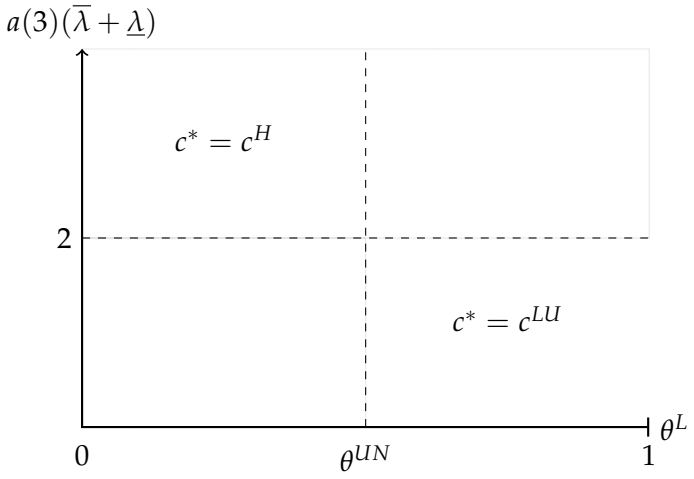


FIGURE 4.4: Optimal Accession Costs - Unanimity

4.4 Discussion and Conclusion

Our paper is a first step towards understanding the interplay of accession terms and voting rules. In a nutshell, we can show that accession costs are only used to extract payments for the incumbents and not as a screening device. To get to this result, we first show that the voting rule is the more efficient screening mechanism to manage the IO's enlargement. Whenever the organization's good production displays large economies of scale, even candidates of low productivity are desirable members and the incumbents are willing to grant veto rights. This entices the candidate to join despite the heterogeneity between them and the founding members. However, if the economies of scale are not large enough to compensate for a potential hold-up problem the incumbents do not give away veto rights and the threat of being overruled in a majority vote is sufficient to deter a low-type candidate from joining. Thus, the IO is perfectly able to screen the desirable candidates by usage of the two voting rules. In the baseline model, accession costs are never put in place because they are the Pareto-inferior screening device.

This finding is of course partly driven by the assumption of accession cost as wasteful spending. Therefore, the second part of the paper models the cost as a transfer from the candidate to the incumbents. We now find that the cost are never set to zero and the incumbents extract as much as possible from the candidate, thereby facing again a trade-off between maximizing the size of the IO, which brings economies of scale, and extracting a larger payment from the high-type candidate, who expects a larger benefit from the IO and thus has a higher willingness-to-pay. Which motive prevails depends mostly on the productivity of the low-type candidate. If they are relatively productive, a lower extraction is set of by the gains in terms of economies of scale. However, if they are less productive, extracting as much as possible from the high-type candidates is the preferred option.

The model is motivated by the example of the Eurozone, which supplies the club good of price stability to all countries using the Euro. Since countries can have opposing views on monetary policy measures in a given period, the simple majority voting with rotating voting rights implies that potential members anticipate that they will not necessarily be able to influence decisions in their favor. Thus, we argue that they join the club only if they are sufficiently similar to the incumbent members, making harsh accession terms obsolete. However, the model is general enough to be applied to any IO that produces a club good.

For a lack of data on accession costs we cannot establish and explain any empirical regularities in the functioning of IOs in practice. Rather, we provide theoretical insights to enable a systematic assessment of the ways in which institutional features can complement and substitute each other. As the analysis focuses on the perspective of the incumbents and how they strategically design institutions in anticipation of other players' behavior in order to maximize their benefit from the IO the contribution of the paper is inherently normative in nature.

The model presented above needs to be assessed in light of its key assumptions. The model is tailored towards IOs in which all members have to set identical policies, e.g. interest rates or tariffs. Whenever this is not critical for the success of the organization, alternative welfare-increasing institutional setups can be considered: for example a coalition of the willing in which low type candidates are allowed to suspend contribution if it is not profitable for them.[6] The free-rider problem – which is central to the majority of literature on international cooperation – is not addressed here. Assuming perfect enforcement is, however, not unrealistic in all the settings where states make payments upfront. This is the case, for instance, in development organizations. Similarly, the EU reports high compliance rates: as of 2017, less than one percent of its directives had not been transposed into national law (European Commission, 2018). The argument can even be extended to claim that countries use IOs as commitment devices to implement unpopular policy measures (see Rotte and Zimmermann (1998) for a discussion of this argument in the EMU context). It thus seems permissible to neglect free-riding in favor of extending previous literature in other aspects. Ultimately, a full generalization of self-enforcing voting rules under endogenous membership would be desirable.

While the model extensions has already relaxed one of the key assumptions, namely that accession payments are made in the form of "burning money", there are several other ways in which the accession cost could be modeled: e.g. as investments into the candidates' productivity or as the result of a bargaining procedure between candidate and IO. The latter approach would allow for more complex equilibria reflecting heterogeneous outside options and candidates' values for the IO. To further briefly discuss the assumption that for all high types and for all low types the same "within-type" draw is realized. The assumption is motivated by the underlying drivers of uncertainty. We argue that the large-scale shocks that motivate our uncertainty would hit all members of a type to the same degree. While this ignores the uncertainty of smaller, more local shocks, our model focuses on the larger-scale shocks.

Further directions of research as outlined here can include the setup of richer models, which generalize to public goods and consequently address the issue of free-riding within the IO. More theoretical work and an empirical investigation are needed to address the issues raised in this paper in a comprehensive way. For the empirical investigation, data on accession costs would be a prerequisite for any kind of large-n study. So far, no dataset on this issue has been compiled.

[6] We thank an anonymous referee for this observation.

Bibliography

Ahrens, Joachim, Herman W Hoen, and Renate Ohr (2005). "Deepening Integration in an Enlarged EU: A Club–theoretical Perspective". In: *Journal of European Integration* 27.4, pp. 417–439.

Ahrens, Joachim, Renate Ohr, and Götz Zeddies (2007). "Enhanced cooperation in an enlarged EU". In: *Review of Economics* 58.2, pp. 130–150.

Alesina, Alberto and Enrico Spolaore (1997). "On the number and size of nations". In: *The Quarterly Journal of Economics* 112.4, pp. 1027–1056.

— (2003). *The Size of Nations*. Cambridge: MIT Press.

Blake, Daniel J and Autumn Lockwood Payton (2015). "Balancing design objectives: analyzing new data on voting rules in intergovernmental organizations". In: *The Review of International Organizations* 10.3, pp. 377–402.

Cancelo, José Ramón, Diego Varela, and José Manuel Sánchez-Santos (2011). "Interest rate setting at the ECB: Individual preferences and collective decision making". In: *Journal of Policy Modeling* 33.6, pp. 804–820.

Dougherty, Keith, Brian Pitts, Justin Moeller, and Robi Ragan (2014). "An experimental study of the efficiency of unanimity rule and majority rule". In: *Public Choice* 158.3-4, pp. 359–382.

Dougherty, Keith L and Julian Edward (2012). "Voting for Pareto optimality: a multidimensional analysis". In: *Public Choice* 151.3-4, pp. 655–678.

Downs, George W, David M Rocke, and Peter N Barsoom (1998). "Managing the evolution of multilateralism". In: *International Organization* 52.2, pp. 397–419.

European Commission (2018). *Single Market Scoreboard 2017 – Transposition*. http://ec.europa.eu/single-market-scoreboard. Accessed: 18.12.2018.

Gilligan, Michael J (2004). "Is there a broader-deeper trade-off in international multilateral agreements?" In: *International Organization* 58.3, pp. 459–484.

Gray, Julia, René Lindstädt, and Jonathan B Slapin (2017). "The dynamics of enlargement in international organizations". In: *International Interactions* 43.4, pp. 619–642.

Hayo, Bernd and Pierre-Guillaume Méon (2013). "Behind closed doors: Revealing the ECB's decision rule". In: *Journal of International Money and Finance* 37, pp. 135–160.

Hooghe, Liesbet and Gary Marks (2015). "Delegation and pooling in international organizations". In: *The Review of International Organizations* 10.3, pp. 305–328.

Irlenbusch, Bernd and Matthias Sutter (2006). "An experimental analysis of voting in the Stability and Growth Pact in the European Monetary Union". In: *Public Choice* 129.3-4, pp. 417–434.

Jonas, Jiri (2006). "Euro adoption and Maastricht criteria: Rules or discretion?" In: *Economic Systems* 30.4, pp. 328–345.

Kandogan, Yener (2000). "Political economy of eastern enlargement of the European Union: Budgetary costs and reforms in voting rules". In: *European Journal of Political Economy* 16.4, pp. 685–705.

Kirchner, Christian (2012). "The unanimity rule revisited: the case of revisions of hybrid constitutions". In: *Public Choice* 152.3-4, pp. 445–449.

Koremenos, Barbara, Charles Lipson, and Duncan Snidal (2001). "The rational design of international institutions". In: *International Organization* 55.4, pp. 761–799.

Kydd, Andrew (2001). "Trust Building, Trust Breaking: The Dilemma of NATO Enlargement". In: *International Organization* 55.4, pp. 801–828.

Maggi, Giovanni and Massimo Morelli (2006). "Self-enforcing voting in international organizations". In: *American Economic Review* 96.4, pp. 1137–1158.

Matsen, Egil and Øistein Røisland (2005). "Interest rate decisions in an asymmetric monetary union". In: *European Journal of Political Economy* 21.2, pp. 365–384.

Moschella, Manuela and Nicola M Diodati (2019). "Does politics drive conflict in central banks' committees? Lifting the veil on the European Central Bank consensus". In: *European Union Politics*, p. 1465116519890412.

Mundell, Robert A (1973a). "A Plan for a European Currency". In: *The Economics of Common Currencies*. Ed. by HG Johnson and AK Swoboda, pp. 143–72.

— (1973b). "Uncommon arguments for common currencies". In: *The Economics of Common Currencies*. Ed. by HG Johnson and AK Swoboda, pp. 114–132.

Posner, Eric A and Alan O Sykes (2014). "Voting Rules in International Organizations". In: *Chicago Journal of International Law* 15.1, p. 195.

Riboni, Alessandro and Francisco J Ruge-Murcia (2010). "Monetary policy by committee: consensus, chairman dominance, or simple majority?" In: *Quarterly Journal of Economics* 125.1, pp. 363–416.

Rotte, Ralph and Klaus F Zimmermann (1998). "Fiscal restraint and the political economy of EMU". In: *Public Choice* 94.3-4, pp. 385–406.

Shackleton and Laffan (1996). "The Budget". In: *Policy Making in the European Union*. Ed. by Wallace H and Wallace W. Oxford: Oxford University Press.

Stone, Randall, Branislav Slantchev, and Tamar London (2008). "Choosing how to cooperate: A repeated public-goods model of international relations". In: *International Studies Quarterly* 52.2, pp. 335–362.

Treaty on the Functioning of the European Union (TFEU) (2012). In: *Official Journal of the European Union*.

Watson, Alison (2004). "Economic and monetary Union: Of currencies and clubs". In: *Journal of European Integration* 26.1, pp. 25–39.

D Appendix to Chapter 4

D.1 Proofs

Proof of Lemma 1

Proof 1 *Consider an IO with unanimity voting. The incumbents' calculus regarding the expected payoff with an additional member is the following:*

$$E[u_{1,2}] = \begin{cases} E[\theta_3^j]a(3)\overline{\lambda} + (1 - E[\theta_3^j]) & \textit{if candidate joins} \\ a(2)\overline{\lambda} & \textit{otherwise} \end{cases}$$

To admit a new member, the difference between the two should be positive. It is always profitable to admit a high type candidate with $\theta^H = 1$ as she inherently votes in line with the incumbents and increases the size of the pie. The net expected utility of the incumbents is strictly positive if a high type candidate joins. Thus, they will never charge accession costs of an amount that would deter high types:

$$c^* < a(3)\overline{\lambda} - 1 \tag{4.7}$$

If a low type enters under unanimity rule, the incumbents reap the benefits of the increased IO size with probability θ^L, but fail to produce at all with probability $(1 - \theta^L)$. Therefore, they prefer to admit also low types if

$$\theta^L a(3)\overline{\lambda} + (1 - \theta^L) - a(2)\overline{\lambda} \geq 0 \iff \theta^L \geq \frac{a(3)\overline{\lambda} - 1}{a(2)\overline{\lambda} - 1} \tag{4.8}$$

Thus, incumbent members are more prone to let new entrants join under unanimity when their probability of a high benefit draw is higher, when the returns to an additional member are larger and when the initial benefits are relatively small.

Thus, the accession cost must be higher than the expected benefits of the low type, but below the expected benefits of the high type. Keep in mind that with unanimity rule the members will receive their outside option of 1 if they do not join the IO at all and – in case they join – if one of the members receives a low benefit draw (since it can veto the production of the good). There is a positive expected benefit if all members receive a high benefit draw. This leads to the following participation constraint for the low type:

$$\theta^L(a(3) \cdot \overline{\lambda} - 1) \leq c \tag{4.9}$$

This gives the lower bound of c. An upper limit is set by the high type's participation constraint:

$$a(3) \cdot \overline{\lambda} - 1 > c \tag{4.10}$$

As long as c simultaneously fulfills these two conditions, a separating equilibrium can be achieved given unanimity voting. Since c is wasteful spending, we can assume the incumbents to choose the social optimum, which is the lowest possible amount and hence

$$c^* = \theta^L(a(3) \cdot \overline{\lambda} - 1) \tag{4.11}$$

.

Proof of Proposition 1

Proof 2 *If a low type expects a positive value of production under simple majority ($\theta^L \geq \theta^*$), such a low type is willing to join a simple majority regime. The candidate is not pivotal and the incumbents therefore benefit from any new member in the IO. The expected gain from participation is larger than the outside option for all players, thus a simple majority rule with no accession cost (see Lemma 1) and everybody joining is an equilibrium.*

If $\theta^{**} < \theta^L < \theta^*$ low types do not join under simple majority, because the expected benefit from membership is lower than the outside option for any c. But the incumbents are strictly better off from letting them join with unanimity and $c = 0$ compared to the IO with two members.

$\theta^L < \theta^{**}$, the candidate is not productive enough to be granted a veto right under unanimity. The incumbents would have to implement $c* = c^{LM}$ to deter entry. Alternatively, they can set the voting rule to simple majority. Since $\theta^L < \theta^*$, low types do not find it profitable to join. Lemma 1 showed that majority rule is always optimally combined with $c = 0$. The incumbents are indifferent between both rule combinations because they produce equivalent effects for the size and function of the IO, but the latter is Pareto-superior, because even while the low candidates join in none of the scenarios, a high type candidate would have to pay $c*$ as well. To maximize welfare simple majority and now accession cost is preferred. We see that any $c > 0$ is never optimal, because given the voting rules as described above, any accession cost is wasteful spending that does not change the size of the organization in equilibrium.

Proof of Lemma 2

Proof 3 *Let us begin by looking at the expected utility from setting the cost equal to c^H, c^{LU} and c^{LM}.*

With $c = c^H$:

$$E[u_{1,2}(c^H)] = \frac{1}{2} \cdot (\frac{1}{2} \cdot c^H + a(3) \cdot \overline{\lambda}) + \frac{1}{2} \cdot a(2) \cdot \overline{\lambda}$$

The first part of the right-hand side is the expected utility from another high-type candidate joining, whereas the second part is the expected utility from a low-type candidate not joining the organisation. This can be rewritten as

$$E[u_{1,2}(c^H)] = \frac{1}{4} \cdot c^H + \frac{1}{2} \cdot [a(3) \cdot \overline{\lambda} + a(2) \cdot \overline{\lambda}]$$

With $c = c^{LU}$ and unanimity rule:

$$E[u_{1,2}(c^{LU})] = \frac{1}{2} \cdot c^{LU} + \frac{1}{2} \cdot a(3) \cdot \overline{\lambda} + \frac{1}{2}[\theta^L \cdot a(3) \cdot \overline{\lambda} + (1 - \theta)]$$

The first term on the right-hand side are the shared accession cost from the new member, the second term is the expected utility from another high-type candidate joining, whereas the third term is the expected utility from a low-type candidate joining the organisation under unanimity rule. This can be rewritten as

$$E[u_{1,2}(c^{LU})] = \theta^L \cdot c^H + \frac{1}{2} \cdot a(3) \cdot \overline{\lambda} + \frac{1}{2}$$

With $c = c^{LM}$ and majority rule:

$$E[u_{1,2}(c^{LM})] = \frac{1}{2} \cdot c^{LM} + a(3) \cdot \overline{\lambda}$$

For simple majority, the expected utility is simply half of the accession cost, since it is split among the incumbents, plus the gains from producing the good. Note that it does not matter whether a low-type or high-type candidate joins, since they will be outvoted by the incumbents anyways.

In a second step, we can identify the values of θ^L below which setting the cost such that the low types will stay out is optimal for each respective voting rule.

Unanimity *Set $c = c^H$ whenever $E[u_{1,2}(c^H)] > E[u_{1,2}(c^{LU})]$, i.e. if and only if the following holds true:*

$$\frac{1}{2} \cdot (\frac{1}{2} \cdot c^H + a(3) \cdot \overline{\lambda}) + \frac{1}{2} \cdot a(2) \cdot \overline{\lambda} > \theta \cdot c^H + \frac{1}{2} \cdot a(3) \cdot \overline{\lambda} + \frac{1}{2}$$

This can be rewritten to

$$\theta^L < \frac{1}{2} \cdot \frac{a(2) \cdot \overline{\lambda} - 1}{a(3) \cdot \overline{\lambda} - 1} + \frac{1}{4}$$

We see that for θ^L lower than $\frac{1}{4}$ the condition will always hold and for θ^L bigger than $\frac{3}{4}$ it will never hold. In between those values the optimal rule depends on the ratio between $a(2)$ and $a(3)$.

Simple Majority *Set $c = c^H$ whenever $E[u_{1,2}(c^H)] > E[u_{1,2}(c^{LM})]$, which is true if and only if the following inequality is satisfied:*

$$\frac{1}{2} \cdot (\frac{1}{2} \cdot c^H + a(3) \cdot \overline{\lambda}) + \frac{1}{2} \cdot a(2) \cdot \overline{\lambda} > \frac{1}{2} \cdot c^{LM} + a(3) \cdot \overline{\lambda}$$

This can be rewritten to

$$\theta^L < \frac{a(2) \cdot \overline{\lambda} - a(3) \cdot (\underline{\lambda} + \frac{1}{2} \cdot \overline{\lambda}) + \frac{1}{2}}{a(3) \cdot (\overline{\lambda} - \underline{\lambda})}$$

Proof of Proposition 2

Proof 4 *To prove the proposition above, we can proceed in three steps for each possible accession cost.*

First, we need to check whether a participation constraint exists for the candidate. This case is only relevant when incumbents want low type candidates to join in a simple majority setting. As shown above (refer to main text), we know that whenever $\theta^L < \theta^*$ low types will not join independent of the cost. Thus, c^{LM} can only be optimal when θ^L is above that threshold.

Second, using Lemma 2, we already know the threshold values of θ^L for which a pooling equilibrium is be optimal.

Third, if θ^L is above all these thresholds, we know that either unanimity with c^{LU} or simple majority with c^{LM} is optimal. To solve for the parameter values for which each rule-cost combination is optimal, we simply need to compare the expected utility from both combinations. More formally, the incumbents will choose simple majority over unanimity if

$$\frac{1}{2} \cdot c^{LM} + a(3) \cdot \overline{\lambda} > \theta \cdot c^H + \frac{1}{2} \cdot a(3) \cdot \overline{\lambda} + \frac{1}{2}$$

This can be rewritten to

$$\theta \cdot (2 - a(3)\overline{\lambda} - a(3)\underline{\lambda}) > 2 - a(3)\overline{\lambda} - a(3)\underline{\lambda}$$

We can see that whenever $2 - a(3)\overline{\lambda} - a(3)\underline{\lambda}$ is smaller than zero, this condition will always hold and if it is larger than zero, it will never hold.

List of Figures

List of Tables