TUM School of Governance TUM School of Management

Munich Papers in Political Economy
Working Paper No. 8/2020

Do rights to resistance discipline the elites? An experiment on the threat of overthrow

Konstantin Chatziathanasiou
Svenja Hippel
Michael Kurschilgen

November 2020

Do rights to resistance discipline the elites?

An experiment on the threat of overthrow*

Konstantin Chatziathanasiou Svenja Hippel Michael Kurschilgen

November 2020

Abstract. The threat of overthrow stabilizes a constitution because it disciplines the elites. This is the main rationale behind rights to resistance. In this paper, we test this conjecture experimentally. We model a society in which players can produce wealth by solving a coordination problem. Coordination is facilitated through a pre-game status-ranking. Compliance with the status hierarchy yields an efficient yet inequitable payoff distribution, in which a player's wealth is determined by her pre-game status. Between treatments, we vary (a) whether overthrows – which reset the status-ranking via collective disobedience – are possible or not, and (b) whether voluntary redistributive transfers – which high-status players can use to appease the low-status players – are available or not. In contrast to established thinking we find that, on average, the threat of overthrow does not have a stabilizing effect as high-status players fail to provide sufficient redistribution to prevent overthrows. However, if an overthrow brings generous players into high-status positions, groups stabilize and prosper. This suggests an alternative rationale for rights to resistance.

Keywords: rights to resistance; civil resistance; constitutional stability; redistribution; coordination; battle of the sexes; experiment

JEL codes: C72, C92, D74, H23, P48

* K. Chatziathanasiou is affiliated to the University of Münster (kchatzia@uni-muenster.de), S. Hippel to the University of Würzburg (svenja.hippel@uni-wuerzburg.de), M. Kurschilgen to the Technical University of Munich and the Max Planck Institute for Research on Collective Goods (m.kurschilgen@tum.de). We thank Stefanie Egidy, Christoph Engel, Jacob Goldin, Oliver Kirchkamp, Julian Nyarko, J.J. Prescott, Cornelius Schneider, Theresa Wenning, as well as seminar participants at MPI Bonn, U Jena, U Maastricht, U Michigan, and the ComplianceNet conference at UC Irvine for valuable comments and discussion. Financial support by the Max Planck Society is gratefully acknowledged.

1 Introduction

Rights to resistance have a rich history in political thought and constitution-making. With such rights, "constitutionmakers give their people a constitutional mandate to overthrow or contradict their governing authorities" (Ginsburg et al. 2012, p. 1184). Typically, rights to resistance legitimize non-violent acts of civil resistance (Roberts and Garton Ash 2009). Occasionally, they even legitimize the use of force (Ginsburg et al. 2012, p. 1227). Rights to resistance are supposed to help stabilize a constitution by reminding the people of their collective power to constrain the rulers. Their inclusion in a constitutional document functions as a precommitment device which disciplines the governing elites (Ginsburg et al. 2012, p. 1184). However, this understanding rests on an important conjecture: that the threat of overthrow does indeed have a disciplining effect. This is not as certain as theory suggests.

Generally, field evidence on the effectiveness of civil resistance has been mixed. The protests of the Arab Spring have figured prominently in the scholarly discussions on both, civil resistance (Roberts et al. 2016) and rights to resistance (Ginsburg et al. 2012, p. 1187). But most of the involved countries have entered enduring phases of conflict and chaos (on the Arab Winter, see The Economist 2016). In contrast, other historic protests had lasting transformative effects (Ackerman and Rodal 2008; Stephan and Chenoweth 2008; Roberts and Garton Ash 2009). A famous example is the Civil Rights movement in the US, which lead to major reforms in the areas of voting, housing, and education (Ackerman 2014). Identifying causal effects of civil resistance in the field is virtually impossible due to endogeneity problems. Therefore, we employ a laboratory experiment to provide causal evidence on a question that is critical: Does the threat of overthrow discipline the elites?

Given a social order in which elites can be distinguished from non-elites, the threat of overthrow could discipline the elites by deterring them from abusing the privileges the social order has endowed them with. We examine this conjecture in a stylized environment. We model a society, in which subjects can produce wealth by solving a coordination problem. Coordination is facilitated through a status-hierarchy, according to which players are ranked. Status is allotted randomly in a pre-game lottery.[1] Solving the coordination problem in accordance with the status-hierarchy is efficient, but has the consequence of inequitable distributional outcomes, as the payoff-hierarchy reproduces the status-hierarchy.

We test the effects of the threat of overthrow by introducing the possibility to reset the ranking via collective disobedience: If, in a given period, the status-hierarchy is disregarded by a critical mass of society, ranks are allotted anew. After the reallotment, low-status players might find themselves

[1] This resonates with the historical fact that exogenous attributes like descent, ethnicity, and gender have played a major role in shaping the access to resources, e.g. land, but also to professions, property rights, and other privileges (Elster 1992; Schotter and Sopher 2003; Moulin 2004). Today, the lifetime income (Bowles and Gintis 2002; Kahlenberg 2010; Chetty et al. 2014; Adermon et al. 2018) but also the political rights (Shachar 2009) of a person still remain heavily influenced by pre-birth factors. Nonetheless, access to resources can be redistributed.

in the positions of former high-status players, and vice-versa. Our overthrow mechanism mirrors resistance through collective non-compliance. While isolated occurrences of non-compliance do not lead to an exchange of elites, this is different if non-compliance is collective and persistent. As social order depends on some level of popular support (for a classical reference, see de La Boétie 1577), collective, large-scale withdrawal of support paired with various forms of protesting behavior, can lead to an exchange of elites (Ackerman and Rodal 2008).[2] How the exchange of elites is achieved *exactly* will vary: It might be that a social movement pushes for elections, it might also be that it puts pressure on parliament to exchange government. In general, achieving an exchange of elites will be onerous, which we mirror in our design.

By reallotting players' positions in the status-hierarchy, recurrent overthrows could – in principle – level wealth over time, albeit with large efficiency losses. In modern societies, leveling of wealth is mainly achieved through redistribution mechanisms, such as taxes and social benefits. For the reduction of disparities, voluntary redistribution can thus be regarded as a substitute to an overthrow. In terms of efficiency, however, redistribution and the threat of overthrow might rather complement one another. In light of a potential overthrow, adequate redistribution might serve to preserve both elite privileges and overall wealth. The idea is exemplified by Acemoglu and Robinson (2000) who understand redistribution measures as strategic investments by elites in order to prevent social unrest and "buy social peace" (Dal Bó and Dal Bó 2011). In this understanding, democratization functions as a precommitment by elites to future redistribution. To study the possible complementarity of overthrow threat and redistribution, we introduce the overthrow mechanism to an environment in which high-status players can make voluntary transfers to low-status players. In this redistributive environment, a credible threat of overthrow should motivate the high-status players to be more generous with their transfers in order to prevent the overthrow from happening. If effective, the threat of overthrow would thus enhance stability by disciplining the greed of the elites.

In contrast to established thinking we find that, on average, the threat of overthrow does not have a stabilizing effect. Higher ranks' willingness to make redistributive transfers does not increase in the presence of a credible overthrow threat. As a consequence, overthrows happen frequently and act as a highly inefficient equalizer. In most experimental groups, high-status players do not react sufficiently to the threat of overthrow. As overthrows happen, these players lose their high status as ranks are reset, and experience a substantial – and preventable – drop in earnings. Some groups, however, experience overthrows only in the early periods of the experiment, and are perfectly stable

[2] Ackerman and Rodal (2008, p. 111) lay out the possible variance of such protesting non-compliance: "Civilians have used disruptive actions as sanctions, to challenge and delegitimate rulers, mobilise publics, constrain authoritarians' power and undermine their sources of support and shift their loyalties. Petitions, marches, walkouts and demonstrations have been used to rouse public support and mobilisation. Forms of non-cooperation such as strikes, boycotts, resignations and civil disobedience have served to frustrate the operations of governments. Direct intervention such as blockades, factory occupations and sit-ins have thwarted rulers' ability to subjugate their people. The sequenced, sustained application of these nonviolent operations has engendered historical results: tyrants have capitulated, governments collapsed, occupying armies retreated and political systems that denied human rights been delegitimated and dismantled."

in later periods. In these stable groups, early overthrows put, eventually, more generous players into high-status positions. By sharing a larger part of their wealth through voluntary redistribution, these players manage to sustain their status. In these stable groups, both the low ranks and the high ranks earn significantly higher payoffs than in the unstable groups. Thus, in the face of an overthrow threat, being generous pays off for high-status players.

Interestingly, we find that the higher transfers in the stable groups are not the result of successful observational learning from the mistakes of earlier (overthrown) high-ranks. Instead, we observe that the high-status players of the stable groups were already generous *before* they became high-status players, and that their generosity correlates with higher levels of trust towards strangers. In sum, while we do not find evidence supporting the notion that overthrows work as a disciplining device, we do find that they can function as a selection mechanism.

The remainder of the paper is structured as follows. Section 2 surveys further related literature. Section 3 introduces the experimental design in detail. Section 4 presents our hypotheses. Our experimental results are presented in Section 5. Our findings are discussed in Section 6.

2 Related Literature

History and Codification of Rights to Resistance. Rights to resist the government have a long tradition. Several historical landmarks stand out. In 1215, the first version of the Magna Carta provided the English nobility with the right to resist the Crown. In 1222, the Golden Bull provided the Hungarian nobility with a right to resist the King. In 1689, John Locke argued for the right to rebel against government, if it breached the People's trust (Locke 1689, § 221 et seq.). In 1776, the *Declaration of Independence of the United States* famously declared the "Right of the People to alter or abolish" government, should it violate the People's rights. In 1789, the French *Declaration of the Rights of Man and of the Citizen* recognized a right to "resistance against oppression". Today, constitutional democracies provide for the right to exchange government through popular vote. At the same time, almost 20% of today's written constitutions contain explicit rights to resistance (Ginsburg et al. 2012, p. 1217). Still, as the case of the Arab Spring illustrates, people do not have to rely on the codification of such a right in order to protest and act against an illegitimate government (Ginsburg et al. 2012, p. 1240). The case also illustrates that rights to resistance do not only aim at ensuring *legality* in a strict sense. They also *legitimize* civil resistance as a political instrument in order to address and pursue social causes (on the political relevance of civil resistance, see Ackerman and Rodal 2008; Stephan and Chenoweth 2008; Roberts and Garton Ash 2009).

Constitutional Stability and Enforcement. Theoretically, rights to resistance play an important role in stabilizing a constitution. Constitutions lack centralized external enforcement (Hardin 1989;

Weingast 1997; Elkins et al. 2009; Hadfield and Weingast 2013). Instead, they have to be enforced decentrally by citizens and officials. Identifying a constitutional transgression and acting against it, requires a coordination effort (Weingast 1997). In such situations, written constitutions and rights function as coordination devices (Elster 1984, p. 95; Hardin 1989, p. 101; Bednar 2009, p. 189; Elkins et al. 2009, p. 7; Hadfield and Weingast 2013). Constitutions provide focal points that enable officials and citizens to identify the limits of government and thus stabilize democracy (Elster 1984, p. 95; Weingast 1997, p. 246). Rights to resistance facilitate this effort. They remind the people of their collective power to constrain the governing authorities (Ginsburg et al. 2012, p. 1209). The inclusion of rights to resistance in a constitutional document are meant to function as a precommitment device that disciplines the behavior of elites. As Ginsburg et al. (2012, p. 1184) note, "the constitutional right to resist can represent a fundamentally democratic and forward-looking tool that constrains future government abuse, empowers the national citizenry, and acts as an insurance policy against undemocratic backsliding". As Ginsburg et al. (2012, p. 1237) point out, it is an open empirical question whether rights to resistance do indeed have such effects.[3] While the possible effects of rights to resistance are manifold[4], our experiment is concerned with the question whether elites can indeed be disciplined via the threat of overthrow.

Inequality and Redistribution. Currently, socio-economic inequality is a major challenge for the stability of constitutions (Sitaraman 2017; Dixon and Suk 2018; Levitsky and Ziblatt 2018; Khaitan 2019). In recent decades, inequality of wealth and income has grown dramatically (Piketty et al. 2014; Milanovic 2016), and law is considered to have contributed to this trend (Pistor 2019). This is problematic as the enforcement of a constitution poses a collective action problem (Hardin 1989; Bednar 2009, p. 170; Ginsburg et al. 2012, p. 1194). Constitutions must be self-enforcing (Weingast 1997; de Figueiredo and Weingast 2005). Officials and citizens must have an incentive to act against a constitutional transgression. High levels of inequality might reduce disadvantaged citizens' motivation to protect the constitutional order. Even worse, some citizens might see it in their interest to support a despotic transgressor (Ryvkin and Semykina 2017). Given these challenges, one possible measure to solidify a constitution can be seen in the establishment of social rights, in order to create a more equitable society (Young 2012; Peters 2017; Moyn 2018). If effective (on doubts, see Chilton and Versteeg 2017), such rights might incentivize citizens to protect the constitution as a whole. But there is also a more cynical approach to social rights. Measured redistribution might serve as a tool

[3] Ginsburg et al. (2012, p. 1237) note that it remains an open question "whether right-to-resist provisions actually make a difference in practice". This implies two sub-questions: first, whether the *codification* of rights to resistance makes a difference, and second, whether rights to resistance are *generally* capable of stabilizing a constitution. Our paper is concerned with the second question.

[4] As Ginsburg et al. (2012, p. 1184) show, the introduction of rights to resist might also be used to *expost* legitimize the actions of coup makers.

to preserve elite privileges (Acemoglu and Robinson 2000) in the face of looming social unrest.[5] Our experiment models a possible link between social inequality and constitutional stability: We create a social order that facilitates societal coordination but benefits people to differing degrees. We then test the effects of potentially stabilizing interventions.

Endogenous Institutions. While we are mainly interested in the *disciplining* effect of rights to resistance and the threat of overthrow, there could potentially be further effects. A range of experiments show that an institution's efficacy is increased if it was generated endogenously, typically through some form of voting mechanism which adds a "democratic dividend" (Tyran and Feld 2006; Dal Bó et al. 2010; Sutter et al. 2010; Markussen et al. 2014; Hauser et al. 2014; DeAngelo et al. 2020; Langenbach and Tausch 2019). The overthrow mechanism in our experiment could add such "democratic dividend", as it enables players to influence the social ordering by recurrently reshuffling the ranks. Moreover, the mere option of overthrow, might already cast a shadow of legitimacy on the status quo. Note, however, that analytically such a legitimizing effect of the threat of overthrow is separate from a disciplining effect on elites. We account for this distinction in our research design by testing the threat of overthrow in two distinct environments: (i) In our baseline environment, elites have no means to react to the overthrow threat. Higher stability would thus necessarily be due to a legitimizing effect. (ii) In our redistributive environment, elites can use voluntary transfers to increase non-elite's participation in the benefits of a stable social order. Higher stability through higher transfer willingness would point toward a disciplining effect.

Focality and Competing Interests. Due to its technical framework, our paper relates to experimental research on coordination games and focal points. Mehta et al. (1994) and Bardsley et al. (2009) show how focal points are effective when players' interests are perfectly aligned. However, Crawford et al. (2008) show that the power of focality is considerably reduced when payoffs are minimally asymmetric. In cases of a very pronounced asymmetry, even explicit recommendations to play a specific equilibrium fail, which leads to substantial efficiency losses (Anbarcı et al. 2018). As one would assume, this is due to the resistance of the players who are asked to play their less preferred equilibrium. The problem can be framed in terms of competing saliencies. A focal point that is salient in the sense of Schelling (1960) competes with the salience of equilibria that yield a higher individual payoff (Crawford et al. 2008, p. 1444). We have taken this research into account in the design of our basic framework of social order. Our framework implements a simple coordination task that can be solved with the help of a focal point. Due to the asymmetry of the payoffs, it can be expected that there will be no full coordination and a substantial welfare loss. This leaves room to observe possible improvement through the behavioral effects of our interventions. At the same

[5] Indeed, while the countries of the Arab Spring differed considerably in many aspects, they all shared high unemployment rates (Roberts 2016, p. 280).

time, our framework does not imply that the social order it creates is legitimate. The existence of a focal point is analytically separate from its legitimacy (Elster 1984, p. 95; Elkins et al. 2007, p. 1146; McAdams 2015, p. 49).

Experimental Regime Change. Our work relates to experimental research on regime change. Ryvkin and Semykina (2017) study a game where citizens can choose to replace a democratic regime with an autocrat. In the democratic regime, property rights are secure and redistribution requires a majority vote, while the autocrat promises full redistribution but can potentially expropriate the citizens. Subjects are more likely to voluntarily switch from democracy to autocracy when inequality is high. This resonates with the aforementioned intuition that inequality might have a destabilizing effect which might benefit autocracy.

Empirical Comparative Law. Our work complements empirical research in comparative constitutional law that is conducted within an explanatory social science framework (Hirschl 2014; Roux 2017). Large-N studies have contributed to our understanding of diverse constitutional phenomena, such as constitutional endurance and change (Elkins et al. 2009; Ginsburg and Melton 2015), diffusion of constitutional provisions (Law and Versteeg 2012; Goderis and Versteeg 2014), adoption of judicial review (Ginsburg and Versteeg 2013), and efficacy of constitutional rights (Chilton and Versteeg 2015, 2017). Yet, while quantitative studies in comparative law can narrow down plausible theories, causal inference is problematic due to endogeneity (Elkins et al. 2009, p. 89; Klick 2010; Spamann 2015). This calls for methodological pluralism (Law 2010; Hirschl 2014, p. 224). Laboratory experiments provide a complement to studies with observational data. They model social phenomena and engage in theory-testing (Schmidt 2009; Falk and Heckman 2009; in the context of law: Camerer and Talley 2007; Engel 2013). While concerns over external validity (Voigt 2011, p. 247) call for a discussion of experimental results (see our Section 6), these concerns pose no general argument against modeling social phenomena and testing the behavioral theories underlying our understanding of institutions.

On a related note, as rights to resistance are not only discussed in comparative law, but also in legal philosophy (see, e.g., Honoré 1988), our paper can also be read as a contribution to the emerging field of experimental jurisprudence (Tobia 2020).

3 Experimental Design

We first present the general framework of our experimental design (3.1), followed by treatments (3.2) and experimental procedure (3.3).

3.1 General Framework

The general framework of our experimental design models a society of N strangers. Over numerous periods, members of this society make random bilateral, anonymous encounters, in which they are faced with a simple coordination problem. Coordination is successful and produces wealth, if the two players who meet in the stage game agree on who *claims* a coveted resource, and who *concedes* it. The stage game is a *Battle-of-the-Sexes (BoS)* as depicted in Figure 1. Due to $h > l > 0$, the game has two pure-strategy Nash equilibria, with each player preferring the equilibrium in which she claims and the other player concedes.

		player 2	
		claim	*concede*
player 1	*claim*	$0, 0$	h, l
	concede	l, h	$0, 0$

Figure 1: Stage Game

The two players who meet in a stage game can solve their coordination problem by following the recommendation of an exogenous coordination device. This mirrors the efficiency enhancing effect of societal coordination through rules, e.g. the aforementioned constitutional rules or rules over property. Our general framework provides for no means to exogenously enforce compliance with the coordination device. The coordination device is provided in the form of a status-ranking, upon which an unambiguous signal is emitted. The status-ranking mirrors a legal or social order that ranks players from highest to lowest degree of privilege. Whenever two players meet, they mutually and unambiguously recognize who is of higher status (and thus supposed to *claim* the resource) and who is of lower status (and thus supposed to *concede* that right to the other player). The higher (lower) one's rank, the more often the action recommended by the coordination device is to claim (concede). The information whether the status of their respective counterpart is higher or lower than their own is all the information that is revealed to them about their counterpart.

There is also a Nash equilibrium in *mixed strategies*. In this *mixed equilibrium*, players ignore the recommendations of the coordination device. Each player chooses her preferred action *claim* with the large probability $\frac{h}{h+l}$ and the less attractive action *concede* with the small probability $\frac{l}{h+l}$. Players thus only generate surplus in the rare occasion that a claiming player meets a conceding player. As a result, there is substantial miscoordination and very little economic surplus. The mixed equilibrium is both individually and socially inefficient, as its expected payoff is below l.[6] This mirrors a situation,

[6] The probabilities of the different outcomes in the mixed equilibrium are: $P(claim, claim) = \frac{h^2}{(h+l)^2}$, $P(claim, concede) = \frac{hl}{(h+l)^2}$, $P(concede, claim) = \frac{hl}{(h+l)^2}$ and $P(concede, concede) = \frac{l^2}{(h+l)^2}$. Players earn zero payoffs whenever the outcome is $(claim, claim)$ or $(concede, concede)$, which happens with probability $\frac{h^2+l^2}{(h+l)^2}$. The mixed equilibrium thus yields an expected payoff of $\frac{hl}{h+l} < l$.

in which society is stuck in an anarchic state, as each player tries to maximize her payoff under the expectation that nobody respects (potentially available) legal institutions.

3.2 *Experimental Treatments*

Based on the general framework, we implement 2×2 treatment variations. We vary (a) whether overthrows – which reset the status-ranking via collective disobedience – are possible or not, and (b) whether voluntary redistributive transfers – which high-status players can use to appease the low-status players – are available or not. Table 1 summarizes our four experimental treatments. In all treatments, players are fully informed about all the parameters and the choice structure of the experimental environment they participate in. For the experimental instructions, please refer to Appendix B.

treatment name	transfer option	overthrow mechanism
baseline	-	-
overthrow	-	✓
transfer	✓	-
transfer-overthrow	✓	✓

Table 1: Experimental Treatments

Baseline. In the *baseline* treatment, the participants are randomly matched into groups of $N = 6$ players. In each period, following a random matching protocol, players within a group are matched in pairs of two and play a BoS as described above. Thus, every period, there are three parallel matches, in which the matched players play a BoS. We choose a parametrization with strong payoff asymmetry and set $h = 10$ and $l = 1$. The actions are color-coded such that players claim by choosing *red*, and concede by choosing *blue*. The parametrization and the actions are depicted in Figure 2.

		player 2	
		red	***blue***
player 1	***red***	0, 0	**10,1**
	blue	1, 10	0, 0

Figure 2: Stage Game Parametrization

In any given stage game, whichever player's letter comes first in the alphabet (as assigned by lottery) sees the action "red" in bold. The other player sees the action "blue" in bold. This is commonly known to all players.

We implement the status-ranking as follows: Before the beginning of the first period, every player in a group of six gets assigned a letter a, b, c, d, e, or f in a pre-game lottery. Every player has the same probability of receiving each letter and every letter gets assigned exactly once. When two players meet, one of the two (pure) stage-game equilibria is highlighted through bold lettering. Which of the two stage-game equilibria is highlighted, depends on whose player's letter comes first in the alphabet. This is commonly known to all players. The exact letter of the other player is never revealed. In any given period, the probability that the advantageous equilibrium will be highlighted equals 1 for type a, 4/5 for type b, 3/5 for type c, and so on. As we have a group of six players that are matched randomly and anonymously in each period without any possibility to communicate, we can rule out that players take turns in playing the more beneficial equilibrium.

Transfer. Whereas in the *baseline* high-status players have only a rather limited action space, in the richer, redistributive environment of the *transfer* treatment, high-status players can additionally share some of their wealth through a voluntary transfer mechanism. Directly after every stage game, players can make a transfer to the counterpart that they just interacted with in that period. The maximum size of a transfer is constrained by the payoff that was generated through successful coordination in this very period, i.e. the player who earned 10 tokens in the stage game, can transfer between 0 and 10 tokens to the other player.[7]

Overthrow and transfer-overthrow. In treatments *overthrow* and *transfer-overthrow*, we introduce an overthrow mechanism to the environments without and with redistribution, respectively. The mechanism functions as follows: If, in the same period, all three matched pairs of a given group fail to coordinate on a payoff-yielding equilibrium, status ranks are allotted anew. The new ranks remain valid until the next overthrow occurs.

Note that triggering the overthrow mechanism is onerous. In expectations, the three lower ranks (d, e, f) would improve their rank after an overthrow while the three higher ranks (a, b, c) would drop in the status ladder. Consequently, the higher ranks have an interest in preventing an overthrow from happening. Even if the three lower-ranked players deliberately disobeyed the coordination device in order to trigger an overthrow, it still might take several periods until the overthrow is achieved. Due to the random matching of pairings in each period, only in 6 of the 15 possible constellations, each of the three lower ranks is matched with one of the three higher ranks (e.g. a with e, and b with d, and c with f). In the remaining 9 constellations, i.e. in 60% of all periods, there is always one pairing of high-status players that – by complying with the recommendation of the coordination device – can prevent an overthrow from happening (e.g. a with b, or a with c, or b with c). Successfully activating the overthrow mechanism thus requires persistence and the willingness to forfeit monetary payoffs.

The overthrow mechanism mirrors resistance to, and overthrow of the governing elites. In the real world, a single act of non-compliance with the status quo may be perceived as a signal of frustration by non-elites to elites, but it will not overthrow government. This changes if non-compliance is collective and persistent. Withdrawal of popular support, typically paired with acts of protest, leads to an exchange of government, i.e. an exchange of elites (for a classical reference, see de La Boétie 1577, who argued that even tyranny hinges on individual support; for modern

[7] Although we expect mainly the high earners to make use of transfers, we also allow for low earners to transfer (up to 1 token) in order to keep our design symmetric. As expected, this almost never happens in the experiment.

examples, see Ackerman and Rodal 2008). How exactly the exchange of elites is achieved, will vary from case to case. It might be that a social movement pushes for elections, it might also be that it puts pressure on parliament to exchange government. In any case, a heavy investment in practices of resistance will be needed. Our experimental design accounts for this stylized fact. Triggering the overthrow mechanism is onerous and inherently inefficient, as it requires players to persistently coordinate on disobeying the status quo.[8]

If the overthrow mechanism is triggered, players again have the same probability to receive a particular letter a, b, c, d, e, or f as in the pre-game lottery. Players can now find themselves in a more privileged position. However, they might also find themselves in a worse position. We thus mirror not only the investment that is necessary to achieve an overthrow but also the risk connected to its realization.

Note that an overthrow (as we model it) does not change the *structure* of the social order. After an overthrow, the players may find themselves in different roles, but the roles *per se* remain the same. Keeping the structure of the social order constant, allows us to draw causal inference from observed behavioral differences between our experimental treatments.[9]

3.3 Procedure

Every subject participated in exactly one experimental treatment (between-subjects design). Each group of six individuals played a total of at least 50 periods. After period 50, the continuation probability dropped to 75%.[10] In each period above 50, there was thus an expected duration of 3 additional periods.[11] All periods were payoff relevant. In the data analysis, we use the first (guaranteed) 50 periods for each individual.

After each period, players received information on the outcomes of their own as well as the other encounters, in the *transfer* and *transfer-overthrow* treatment they then decided on a transfer, and finally received the list of total payoffs. In case the overthrow mechanism was triggered, participants were informed of their new status rank, just before proceeding to a new period.

The instructions were handed out in print to the subjects (see Appendix B) and read aloud, before the experiment started. To ensure full understanding, subjects had to pass extensive control questions. We implemented the experiment with as little contextual information as possible. While contextualization of decisions can sometimes be useful in experiments (Alekseev et al. 2017), decontextualization allows for better experimental control (Smith 1976, p. 278; on limitations, see Engel

[8] Note that in the real world, it would also be possible to free-ride on the revolutionary efforts of other members of a society. In our experiment, such free-riding is only possible in a weak sense. Players are informed of occuring miscoordinations which could be interpreted as a signal that overthrow is more likely. While it is possible to free-ride on the provision of such a signal, it is not possible to free-ride on the actual overthrow, as the latter requires that all matches miscoordinate.

[9] This resonates with the fact that the actions which are typically protected by rights to resistance aim at restoring a constitutional order, not at replacing one order with a different one (Ginsburg et al. 2012, p. 1193).

[10] This procedure closely follows the protocol of Camera and Casari (2009).

[11] $.75 + .75^2 + .75^3 + .75^4 + .75^5 + ... = 3$.

and Rand 2014). This is of particular importance in relation to experiments on institutions, where experimental subjects are likely to have rich preconceptions that may affect their choices (McAdams 2015, p. 63).[12]

After the game ended, we elicited participants' (a) other-regarding preferences (Liebrand and McClintock 1988) and (b) risk and trust attitudes as commonly elicited in the German Socioeconomic Panel (SOEP) as well as (c) some socio-demographics (age, gender, number of siblings). For more details on the post-experimental tests, see Appendix C.

The present paper builds on our previously published research. For the treatments *baseline* and *transfer*, we report data from Chatziathanasiou et al. (2020). Data for *overthrow* and *transfer-overthrow* is being reported for the first time. All data was collected in one wave. The experiment was conducted at the BonnEconLab of the University of Bonn, Germany, and computerized using the software z-Tree (Fischbacher 2007). From a database of more than 5000 people, we recruited 378 subjects[13] using hroot (Bock et al. 2014). Subjects were mainly undergraduate students from a variety of disciplines. Sessions lasted about 90 minutes and subjects earned on average 16.47 € (about 19.38 $), including a show-up fee of 4 €. During the experiment, payoffs where presented in experimental currency units (ECU), with a known exchange rate of ECU 25 = 1 €. Subjects sat in visually completely isolated cubicles.

4 Hypotheses

This section develops and presents our hypotheses. In order to discipline the elites and benefit stability, the threat of overthrow has to satisfy two conditions. First, the threat needs to be *credible*, i.e. it needs to have a high chance of materializing if it is not actively prevented. Second, the elites need to *adapt* to the threat, i.e. they have to do what is necessary to prevent the overthrow from happening.

Credible threat. The threat of overthrow can be considered credible if the introduction of the overthrow mechanism comes with a substantial increase in miscoordination, and thus with a high likelihood of an actual overthrow being triggered. The main reason for expecting an increase in miscordination is the changed incentive structure for the low-status players when overthrows are possible. For low-status players, the expected payoff after a reallotment of ranks is (weakly) higher than the expected payoff without a reallotment.

For a given period, the expected payoff of every rank depends on the probability to meet a

[12] Preconceptions about law and legitimacy can themselves be a matter of experimental investigation (see, e.g., Engel and Kurschilgen 2013; Chilton 2014). However, in this experiment, for achieving higher levels of control and generality, we deliberately avoid invoking subjects' perceptions of existing institutions.

[13] 96 subjects in the treatments *baseline*, *transfer* and *overthrow*, respectively, and 90 subjects in *transfer-overthrow*. Over 50 periods, we thus have a total of 4800 (or 4500) observations per treatment, clustered in 16 (or 15) statistically independent groups per treatment.

lower-ranked or higher-ranked player, and on that player's propensity to follow the coordination device. For rank a the probability to meet a lower-ranked player and achieve the high payoff of $h = 10$ is 100%. Denoting the probability that the other player follows the coordination device by $\lambda \in [0,1]$, player a expects a payoff of 10λ tokens in a given period. Similarly, the payoffs of the other ranks are 8.2λ for player b, 6.4λ for player c, 4.6λ for player d, 2.8λ for player e, and 1λ for the lowest-rank player f.[14] The *average rank* has expected payoffs of 5.5λ tokens. If an overthrow occurs and ranks are reset, every player has the same probability to be allotted a certain rank. In expectations, each player becomes an *average rank* after the reset. From the perspective of ranks d, e and f, becoming an *average rank* (who earns 5.5λ in a given period) is more attractive than conserving their current rank. When overthrows are possible, the lower ranks have an incentive to trigger an overthrow, by avoiding coordination on the Nash equilibrium recommended by the coordination device.

Potentially, players' risk aversion could mitigate this incentive to miscoordinate. Even for the low-status players there are risks connected to an overthrow. While the low-status players earn at least some small amounts when following the coordination device, these gains might be lost if society slides into an anarchic state. In particular, other players' propensity to follow the coordination device (λ) may drop in the wake of an overthrow, which would reduce the expected payoff after an overthrow and thus its appeal. The erosion of players' willingness to follow the device could eventually lead to the highly unattractive mixed equilibrium, in which all players ignore the recommendations of the coordination device and have expected payoffs of $\frac{10}{11}$ tokens. But as long as risk aversion does not completely nullify the miscoordination incentive (which is unlikely), the possibility of overthrow should decrease lower-ranks' willingness to follow the coordination device. Thus, we formulate the following hypothesis:

Hypothesis 1. *In the absence of transfer opportunities, the introduction of the overthrow mechanism will lead to an increase in miscoordination.*

Adaptation of elites. In view of a credible overthrow threat, elites have more to lose. In the presence of transfer opportunities, they could adapt to this threat by displaying a higher willingness to redistribute economic surplus to lower ranks. Voluntary transfers have a soothing function, as they reduce the gap between the payoffs of low-status players and high-status players (Chatziathanasiou et al. 2020). From a theoretical perspective, voluntary redistribution can be understood as a way of honoring a tacit social contract (Binmore 1998). In exchange for appropriate compensation, low-status players accept the status quo and concede privilege to whomever has higher status. If, however, high-status players do not make sufficient transfers, low-status players now have an indirect

[14] The general formula for the expected payoff in one period given coordination according to the device is: $\frac{n-1}{N-1} l\lambda + \frac{N-n}{N-1} h\lambda$ with N being the number of players in the society and n being the numerical rank of a player, i.e. $n = 1$ for rank a, $n = 2$ for rank b, etc.

punishment tool at their disposal. This should discipline high-status players and motivate them to make higher transfers. In turn, higher transfers should lead to more compliance with the status quo by the low-status players. Overall we should thus observe a higher level of efficiency, as higher coordination leads to higher wealth generation. This leads us to the following hypotheses:

Hypothesis 2. *High-status players will adapt their behavior to the threat of overthrow and display a higher willingness to transfer.*

Hypothesis 3. *In the presence of transfer opportunities, the threat of overthrow will increase compliance with the status quo and lead to higher levels of efficiency.*

5 Results

We first report treatment differences on the incidence of overthrow events, and then look into payoff implications. Subsequently, we will zoom into the *transfer-overthrow* treatment and examine the behavior of high-status players in the face of potential overthrow in order to understand why certain groups manage to stabilize while others do not.[15]

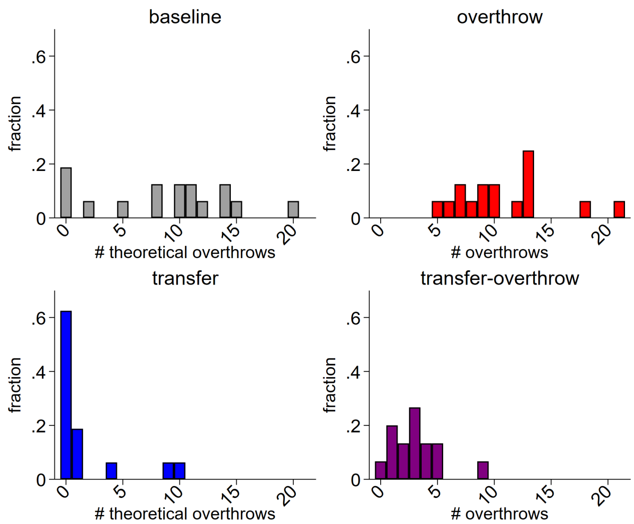


Figure 3: Incidence of Overthrow Events

Distribution of the number of overthrow events, by treatment. An overthrow event is a situation in which all matched pairs fail to coordinate on a payoff-yielding equilibrium in a given period.

Overthrow Events. Figure 3 shows the number and distribution of overthrow events across treatments. We denote as *overthrow event* a situation in which all matched pairs fail to coordinate on a payoff-yielding equilibrium in a given period. As noted in Section 3, the miscoordination of all three pairings triggered an overthrow of the prevailing status hierarchy in treatments *overthrow* and *transfer-overthrow* whereas for *baseline* and *transfer* the overthrow event is purely theoretical.

[15] Unless stated differently, for comparisons between treatments we report two-sided Wilcoxon rank-sum tests over group means.

In the absence of transfer opportunities, overthrow events are rather frequent. In *baseline*, only 19% of all groups never experience an overthrow event. The mere possibility of overthrow makes overthrow events, i.e. full miscoordination in a given period, even more likely ($p = 0.073$). In the *overthrow* treatment there is not a single group that never experiences an overthrow. Put differently, in the absence of transfers, the probability that a group experiences at least one overthrow, and thus the status hierarchy is reset, equals 100%. Thus, the threat of an overthrow is credible and the first condition for a disciplining effect is fulfilled, yielding support for Hypothesis 1:

Result 1. *In the absence of redistribution, overthrows occur in all groups.*

In principle, appropriate redistributive transfers could prevent overthrows from happening. And indeed, over the course of 50 periods the average group experiences 8.8 (hypothetical) overthrow events in *baseline* but only 1.6 (hypothetical) overthrow events in *transfer* ($p < 0.001$). Similarly, the average number of (real) overthrows drops from 10.9 in *overthrow* to 3.1 in *transfer-overthrow* ($p < 0.001$).

Strikingly, however, in the presence of transfer opportunities we do not find overthrow events to become less likely when the threat of overthrow is real. Contrary to the disciplining conjecture, they actually become *more* likely. Whereas in *transfer*, the share of groups that never experience a single overthrow event is 63%, in *transfer-overthrow* this share drops to merely 7% ($p < 0.001$). Put differently, in 93% of groups in *transfer-overthrow* the initial high ranks were unable to sustain the status order and consequently lost their high status.

Result 2. *On average, overthrows do not have a stabilizing effect. Despite redistribution opportunities, only 7% of initial high ranks manage to preserve their initial high status.*

Payoff Implications. Figure 4 compares the payoff development over time in all treatments. We distinguish between players who were allotted a high rank (a, b, c) at the beginning of the game, and those who were initially allotted a low rank (d, e, f). In the absence of transfer opportunities, the overthrow mechanism acts as an equalizer, leveling the differences in payoffs among the players. Initial lower ranks earn significantly more in *overthrow* than in *baseline* ($p = 0.006$) whereas initial upper ranks earn significantly less when overthrows are possible ($p = 0.004$). The Gini-coefficient drops significantly from .30 in *baseline* to .18 in *overthrow* ($p < 0.001$).

The overthrow mechanism still acts as an equalizer in the presence of transfer opportunities. Despite the fact that, in principle, voluntary transfers would be a more efficient substitute for that purpose. The Gini-coefficient drops from .18 in *transfer* to .15 in *transfer-overthrow* ($p = 0.058$). Underscoring the inefficiency, the substantial payoff drop of the initial upper ranks in *transfer-overthrow* ($p = 0.010$) is not matched by a payoff increase of the initial lower ranks ($p = 0.441$). As illstrated in the right-hand panel of Figure 4, the payoffs of the initial upper ranks start off at a

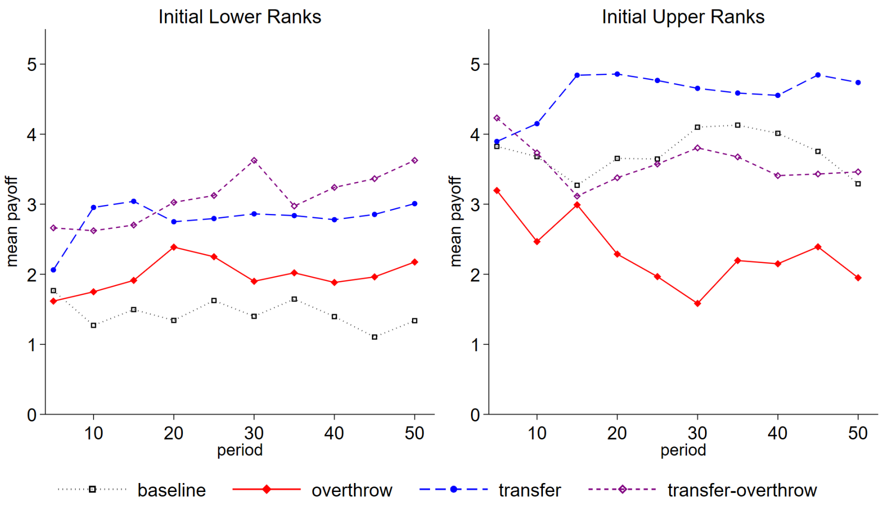


Figure 4: Payoffs over Time by Initial Rank

Mean payoffs (after transfers) over time. We distinguish between players who were allotted a high rank (a, b, c) at the beginning of the game, and those who were initially allotted a low rank (d, e, f).

similar level, around 4, in both treatments but then diverge over time. While they increase to about 4.9 in *transfer* as coordination improves, they decrease to about 3.4 in *transfer-overthrow* as overthrows occur and initial high ranks lose their privileged position.

As stated in Section 4, the second condition for a disciplining effect requires high status players to adapt to the threat of being overthrown. In the presence of redistribution opportunities, high status players could, in principle, react by increasing their transfer levels in order to appease the low ranks and prevent overthrows. But the level of transfers in *transfer* (1.99) and *transfer-overthrow* (2.06) is almost identical (p = 0.843).[16] Thus, we find no support for Hypothesis 2 and note:

Result 3. *High-status players fail to adapt their behavior to the threat of overthrow, despite having more to lose.*

As a result, the threat of overthrow fails to have a positive impact on overall efficiency (see Figure A2). In the absence of transfer opportunities, groups generate less than half of the payoffs that would be attainable if all players complied with the recommendation of the coordination device: 47% in *baseline* and a mere 39% in *overthrow*. In the presence of transfer opportunities, the threat of overthrow actually reduces efficiency (weakly) from 67% in *transfer* to 61% in *transfer-overthrow* (p = 0.093). Thus, we reject Hypothesis 3 and note:

Result 4. *The threat of overthrow does not lead to higher compliance with the status quo, nor to higher efficiency.*

[16] See Figure A1.

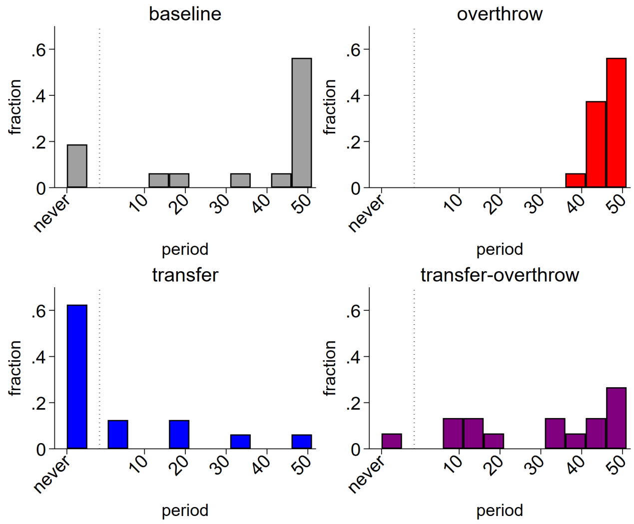


Figure 5: Period of Last Overthrow Event

Distribution of the period of the last overthrow event in a given group, by treatment. An overthrow event is a situation in which all matched pairs fail to coordinate on a payoff-yielding equilibrium in a given period.

Stable vs. Unstable Groups. When overthrows occur, high rank players lose their status as ranks are reset, and experience a substantial drop in earnings. Almost all groups in *overthrow* and *transfer-overthrow* experience (one or more) overthrows in the first half of the game (see Figure A3). In the absence of transfers, all groups continue to experience overthrow events throughout the game. But in the presence of transfer opportunities, some groups become perfectly stable in the second half of the game. Figure 5 shows the period in which an overthrow occurs *for the last time* in a given group. In the following, we explore the differences between the 6 groups from the *transfer-overthrow* treatment that do not experience an overthrow after period 25 (*stable groups*) and the 9 groups that experience one or more overthrows after period 25 (*unstable groups*).

What distinguishes stable groups from unstable groups in the *transfer-overthrow* treatment? Two plausible mechanisms come to mind. On the one hand, stable groups could simply be the result of successful observational learning from the mistakes of earlier high-ranks. Initial low ranks may observe that transfers of the initial high ranks are too low to prevent overthrows. When eventually put into a high-status position in the aftermath of an overthrow, those players start giving more generous transfers for *strategic* reasons, i.e. in order to preserve their status. On the other hand, the stable groups may just benefit from the fact that overthrows put, eventually, more *intrinsically* generous players into high-status positions.

To shed light into the mechanism, we take advantage of the fact that (a) the group formation, (b) the allotment (and reallotment) of ranks within a group, and (c) the pairing of players in a given period were exogenous. In addition, our design allows us to not only observe the generosity of high ranks (a, b, c) but also of low ranks (d, e, f). Whenever a player earned 10 tokens in the stage game, she could transfer up to 10 tokens to the other player. The lower a player's rank, the less frequently she earned 10 (because she was usually matched with higher-ranked players), but from time to time lower ranks did earn 10, and could thus display their generosity.

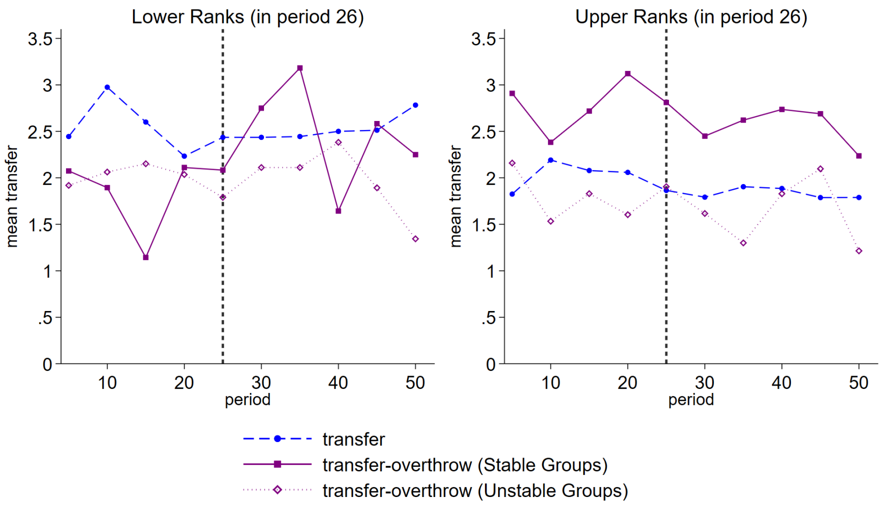


Figure 6: Stability and Transfers

Mean transfers. 6 of 15 groups in *transfer-overthrow* are stable (i.e. no overthrow after period 25).

Figure 6 shows average transfers over time, distinguishing between groups that were stable (i.e. experienced no overthrow in periods 26-50) or unstable in the second half of the game, and between players who were high ranks or low ranks at the beginning of the second half (i.e. in period 26). For the lower ranks, there are no systematic differences. But for the upper ranks, there are two important differences. First, in the second half of the game, high ranks in stable groups are significantly (p = 0.022) more generous (transferring on average 2.5 tokens) than high ranks in unstable groups (1.4 tokens). Stability thus emerges in groups with systematically higher redistribution. Second, the high-status players of the stable groups did not learn to become more generous by observing the failure of the initial high ranks. Rather, they were already more generous *before* they became high-status players. In the early periods of the game, the later-to-be high-status players of the stable groups transfer on average 2.7 tokens whereas the later-to-be high-status players of the unstable groups transfer on average 1.8 tokens (p = 0.039).

These findings are confirmed by regression analysis. Column 1 of Table 2 shows that the probability of a group becoming stable in the second half of the game increases by 10.3 percentage points (p = 0.030) when a player who is of high-status in period 26 gave on average 1 token more in the first half of the game. Column 2 shows that a player's generosity early in the game predicts her generosity late in the game. A player who gives 1 token more in the first half gives, on average, .755 tokens more in the second half (p < 0.001). Column 3 shows that generosity is mainly associated with higher levels of individual trust towards strangers (p = 0.058) measured in the post-experimental tests. Indeed, trust has been shown to be a critical ingredient of economic development (Algan and Cahuc 2010; Bigoni et al. 2016) and the stability of legal institutions (Licht et al. 2005; Alesina and

Giuliano 2015).

Result 5. *The new elites do not learn from the mistakes of the old elites. Stability emerges when intrinsically generous players are (accidentally) allotted high-status positions.*

Table 2: Determinants of Stable Groups

	(1) DV: *stable group*	(2) DV: *second-half transfers*	(3) DV: *first-half transfers*
first-half transfers	0.103** (0.05)	0.755*** (0.13)	
ringdegree			0.589 (0.46)
trust			1.511* (0.73)
risk			-0.336 (0.36)
gender			-0.267 (0.26)
siblings			0.186 (0.17)
age			0.0129 (0.04)
Constant		0.184 (0.22)	2.431*** (0.43)
Observations	45	45	45
Clusters	15	15	15
R^2		0.577	0.132

Column (1) shows marginal effects of probit regressions. Columns (2) and (3) show coefficients of OLS regressions. Cluster-robust standard errors in parentheses. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Who benefits from stability? Figure 7 shows average payoffs over time, distinguishing between stable or unstable groups, and between players who were high rank or low rank at the beginning of the second half (i.e. in period 26). We find that, unsurprisingly, being in a stable group is beneficial for the lower ranks as they receive higher transfers. After period 25, low-status players in the stable groups have significantly higher earnings (unstable 2.66; stable 3.35; p = 0.046). Strikingly, however, being generous pays off for the high-status players, too. By sharing a larger part of their wealth through voluntary redistribution, the high ranks of stable groups not only manage to sustain their status, they actually end up earning significantly more than the high rank players in the unstable groups. After period 25, the mean payoff for high-status players in a stable group is 5.01, as compared to 3.31 in an unstable group (p < 0.001).

Result 6. *High status players' generosity leads to stability and higher payoffs both for low-status and for high-status players.*

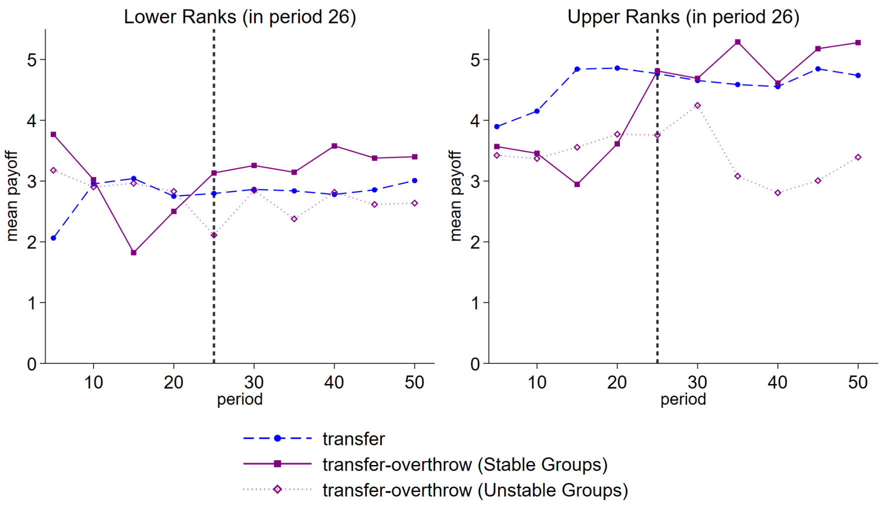


Figure 7: Stability and Payoffs

Mean payoffs after transfers. 6 of 15 groups in *transfer-overthrow* are stable (i.e. no overthrow after period 25).

6 Discussion and Outlook

Despite being credible and costly, the threat of overthrow did not, on average, have a disciplining effect on the elites. In most experimental groups, high-status players did not adapt their transfer behavior to the threat of overthrow. As a result, overthrows happened and the initial elites lost their privileged position. However, in the second half of the game some groups stabilized and prospered. In these groups, the overthrow mechanism brought – eventually, through the reshuffling of ranks – players into high-status positions who were intrinsically generous. The price of stability was not too high for those generous high ranks. To the contrary, they earned significantly more than the (less generous) high ranks of the unstable groups.

Stylized laboratory experiments do not fully map the rich context of the real world. They model and test behavioral mechanisms and theories. Our experiment modeled redistributive choices and reactions to them. A solid body of research (Sausgruber and Tyran 2011; Esarey et al. 2012; Durante et al. 2014) has documented that redistributive choices in the laboratory have predictive power for the real world. Revealed preferences elicited in the laboratory are largely in line with observational field data (Fong 2001; Alesina and Angeletos 2005; Alesina and Giuliano 2011), not only of the general population but also of the economic elite Cohn et al. (2019). In a recent study in Switzerland, laboratory behavior has been shown to predict individuals' support for redistributive policies in national plebiscites (Epper et al. 2020).

What are the implications of our findings? In comparative law scholarship, rights to resistance have been understood to be a precommitment device that disciplines the elites. But our experiment shows that elites may be too myopic to adapt to the threat of overthrow, even if this threat is real and credible. In our experiment, players had full discretion to use redistribution as a means to prevent overthrows. But, on average, elites failed to exercise their discretion to that end. Based on this failure, one could make an argument for reducing discretion. In the real world, one possible instrument to reduce discretion over redistribution would be the constitutionalization of wealth-sharing policies, for instance in the form of socio-economic rights. Admittedly, this would then lead to the question whether socio-economic rights are effective (see Chilton and Versteeg 2017).

But our findings also suggest an alternative rationale for rights to resistance. In the experimental groups that ended up being stable and prosperous, overthrows did not function as a disciplining device, but instead as a selection mechanism. In those groups, elites where exchanged via the overthrow mechanism until eventually, generous players were thrust into high-status positions. In the real world, old elites can resort to authoritative means to prevent such change. In our experiment, this was excluded by design. On the one hand, this only underlines the myopia of the old elites. In a setting with no coercive means, elites should be even more sensitive to the threat of overthrow. On the other hand, the finding that – absent coercion – overthrow functions as a screening device resonates with real world experiences. Among the countries of the Arab Spring, only the case of Tunesia is considered a success (Willis 2016; Roberts 2016). In Tunesia, persistent and recurring protests continued even after the ousting of former President Ben Ali. Subsequent governments agreed to peaceful transitions of power. A nuanced description of the case points to an "attachment to values of procedure, process, and forms of legality [...] that marks out Tunesia and explains the success of forms of civil resistance" (Willis 2016, p. 52). In 2015, the reform-oriented negotiation parties were awarded the Nobel Peace Prize (The Norwegian Nobel Committee 2015). The case illustrates the potential of persistent protests: They shaped the compositions of subsequent governments, until – eventually – a reform-oriented constellation was found.

References

Acemoglu, D. and J. A. Robinson (2000). Why did the west extend the franchise? Democracy, inequality, and growth in gistorical perspective. *The Quarterly Journal of Economics 115*(4), 1167–1199.

Ackerman, B. (2014). *We the People: The Civil Rights Era*. Cambridge: Harvard. University Press.

Ackerman, P. and B. Rodal (2008). The strategic dimensions of civil resistance. *Survival. Global Politics and Strategy 50*(3), 111–126.

Adermon, A., M. Lindahl, and D. Waldenström (2018). Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations. *The Economic Journal 128*(612), F482–F513.

Alekseev, A., G. Charness, and U. Gneezy (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization 134*, 48–59.

Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *The American Economic Review 95*(4), 960–980.

Alesina, A. and P. Giuliano (2011). Preferences for redistribution. In *Handbook of Social Economics*, Volume 1, pp. 93–131. Elsevier.

Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of Economic Literature 53*(4), 898–944.

Algan, Y. and P. Cahuc (2010). Inherited trust and growth. *American Economic Review 100*(5), 2060–92.

Anbarcı, N., N. Feltovich, and M. Y. Gürdal (2018). Payoff inequity reduces the effectiveness of correlated-equilibrium recommendations. *European Economic Review 108*, 172–190.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2009). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal 120*(543), 40–79.

Bednar, J. (2009). *The Robust Federation. Principles of Design*. Cambridge University Press.

Bigoni, M., S. Bortolotti, M. Casari, D. Gambetta, and F. Pancotto (2016). Amoral familism, social capital, or trust? the behavioural foundations of the italian north–south divide. *The Economic Journal 126*(594), 1318–1341.

Binmore, K. (1998). *Game Theory and the Social Contract: Just Playing (Vol. 2)*. MIT Press.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*, 117–120.

Bowles, S. and H. Gintis (2002). The inheritance of inequality. *Journal of Economic Perspectives 16*(3), 3–30.

Camera, G. and M. Casari (2009). Cooperation among strangers under the shadow of the future. *The American Economic Review 99*(3), 979–1005.

Camerer, C. and E. Talley (2007). Experimental study of law. In A. M. Polinsky and S. Shavell (Eds.), *Handbook of Law and Economics, Vol. 2*, pp. 1619–1650. Elsevier.

Chatziathanasiou, K., S. Hippel, and M. Kurschilgen (2020). Property, redistribution, and the status quo – A laboratory study. *Experimental Economics*, Forthcoming.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? The geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics 129*(4), 1553–1623.

Chilton, A. S. (2014). The influence of international human rights agreements on public opinion: An experimental study. *Chicago Journal of International Law 15*, 110–137.

Chilton, A. S. and M. Versteeg (2015). The failure of constitutional torture prohibitions. *The Journal of Legal Studies 44*(2), 417–452.

Chilton, A. S. and M. Versteeg (2017). Rights without resources: the impact of constitutional social rights on social spending. *The Journal of Law and Economics 60*(4), 713–748.

Cohn, A., L. Jessen, M. Klasnja, and P. Smeets (2019). Why do the rich oppose redistribution? An experiment with America's top 5%. *mimeo*.

Crawford, V., U. Gneezy, and Y. Rottenstreich (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *The American Economic Review*, 1443–1458.

Dal Bó, E. and P. Dal Bó (2011). Workers, warriors, and criminals: Social conflict in general equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Dal Bó, P., A. Foster, and L. Putterman (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *The American Economic Review 100*(5), 2205–2230.

de Figueiredo, Rui J. P., J. and B. R. Weingast (2005, 04). Self-Enforcing Federalism. *The Journal of Law, Economics, and Organization 21*(1), 103–135.

de La Boétie, É. (1577). *Discours de la servitude volontaire* (The Politics of Obedience: The Discourse of Voluntary Servitude. 1975, translated by Harry Kurz ed.). Free Life Editions.

DeAngelo, G. J., D. Dubois, and R. Romaniuc (2020). The perils of democracy. *Journal of Economic Behavior & Organization 175*, 328–340.

Dixon, R. and J. Suk (2018). Liberal constitutionalism and economic inequality. *University of Chicago Law Review 85*, 369–401.

Durante, R., L. Putterman, and J. van der Weele (2014). Preferences for redistribution and rerception of fairness: An experimental study. *Journal of the European Economic Association 12*(4), 1059–1086.

Elkins, Z., T. Ginsburg, and J. Melton (2007). Baghdad, Tokyo, Kabul...: Constitution making in occupied states. *William & Mary Law Review 49*, 1139–1178.

Elkins, Z., T. Ginsburg, and J. Melton (2009). *The Endurance of National Constitutions*. Cambridge University Press.

Elster, J. (1984). *Ulysses and the Sirens. Studies in Rationality and Irrationality*. Cambridge University Press.

Elster, J. (1992). *Local justice: How institutions allocate scarce goods and necessary burdens*. Russell Sage Foundation.

Engel, C. (2013). *Legal Experiments: Mission Impossible?* Eleven International Publishing.

Engel, C. and M. Kurschilgen (2013). The coevolution of behavior and normative expectations: an experiment. *American law and economics review 15*(2), 578–609.

Engel, C. and D. G. Rand (2014). What does "clean" really mean? The implicit framing of decontextualized experiments. *Economics Letters 122*(3), 386–389.

Epper, T., E. Fehr, and J. Senn (2020). Other-regarding preferences and redistributive politics. *mimeo*.

Esarey, J., T. C. Salmon, and C. Barrilleaux (2012). What motivates politicalpreferences? Self-interest, ideology, and fairness in a laboratory democracy. *Economic Inquiry 50*(3), 604–624.

Falk, A. and J. J. Heckman (2009). Lab experiments are a major source of knowledge in the social sciences. *Science 326*(5952), 535–538.

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

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public economics 82*(2), 225–246.

Ginsburg, T., D. Lansberg-Rodriguez, and M. Versteeg (2012). When to overthrow your government: The right to resist in the world's constitutions. *UCLA L. Rev. 60*, 1184–1260.

Ginsburg, T. and J. Melton (2015). Does the constitutional amendment rule matter at all? Amendment cultures and the challenges of measuring amendment difficulty. *International Journal of Constitutional Law 13*(3), 686–713.

Ginsburg, T. and M. Versteeg (2013, 07). Why Do Countries Adopt Constitutional Review? *The Journal of Law, Economics, and Organization 30*(3), 587–622.

Goderis, B. and M. Versteeg (2014). The diffusion of constitutional rights. *International Review of Law and Economics 39*, 1–19.

Hadfield, G. and B. Weingast (2013). *Constitutions as Coordinating Devices. USC CLEO Research Paper C11-20*.

Hardin, R. (1989). Why a constitution? In B. Grofman and D. Wittman (Eds.), *The Federalist Papers and the New Institutionalism*, pp. 100–120. Agathon Press.

Hauser, O., D. Rand, A. Peysakhovich, and M. Nowak (2014). Cooperating with the future. *Nature 511*(7508), 220–223.

Hirschl, R. (2014). *Comparative Matters. The Renaissance of Comparative Constitutional Law*. Oxford University Press.

Honoré, T. (1988). The right to rebel. *Oxford Journal of Legal Studies 8*, 34–54.

Kahlenberg, R. D. (2010). *Affirmative action for the rich: Legacy preferences in college admissions*. Century Foundation Press New York, NY.

Khaitan, T. (2019). Political insurance for the (relative) poor: How liberal constitutionalism could resist plutocracy. *Global Constitutionalism 8*(3), 536–570.

Klick, J. (2010). The perils of empirical work on institutions: Comment. *Journal of Institutional and Theoretical Economics 166*(1), 166–170.

Langenbach, P. and F. Tausch (2019). Inherited institutions: Cooperation in the light of democratic legitimacy. *The Journal of Law, Economics, and Organization 35*(2), 364–393.

Law, D. (2010). Constitutions. In P. Cane and H. Kritzer (Eds.), *The Oxford Handbook of Empirical Legal Research*, pp. 376–398. Oxford University Press.

Law, D. and M. Versteeg (2012). The declining influence of the united states constitution. *NYUL Rev. 87*, 762–858.

Levitsky, S. and D. Ziblatt (2018). *How Democracies Die*. Crown.

Licht, A. N., C. Goldschmidt, and S. H. Schwartz (2005). Culture, law, and corporate governance. *International review of law and economics 25*(2), 229–255.

Liebrand, W. and C. McClintock (1988). The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality 2*(3), 217–230.

Locke, J. (1689). Two treatises of government. In P. Laslett (Ed.), *Locke: Two Treatises of Government* (1988, Student ed.)., Cambridge Texts in the History of Political Thought. Cambridge University Press.

Markussen, T., L. Putterman, and J.-R. Tyran (2014). Self-organization for collective action: An experimental study of voting on sanction regimes. *Review of Economic Studies 81*(1), 301–324.

McAdams, R. (2015). *The Expressive Powers of Law: Theories and Limits*. Harvard University Press.

Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review 84*(3), 658–673.

Milanovic, B. (2016). *Global Inequality*. Harvard University Press.

Moulin, H. (2004). *Fair Division and Collective Welfare*. MIT Press.

Moyn, S. (2018). *Not enough: Human rights in an unequal world*. Harvard University Press.

Peters, A. (2017). *Global Constitutionalism: The Social Dimension. MIPL Research Paper Series, No. 2017-25*. Heidelberg: Max Planck Institute for Comparative Public Law and International Law.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy 6*(1), 230–271.

Pistor, K. (2019). *The code of capital: How the law creates wealth and inequality*. Princeton University Press.

Roberts, A. (2016). Revolt for dignity, tunisia's revolution and civil resistance. In A. Roberts, M. J. Willis, R. McCarthy, and T. G. Ash (Eds.), *Civil resistance in the Arab Spring: Triumphs and disasters*, pp. 271–325. Oxford University Press.

Roberts, A. and T. Garton Ash (2009). *Civil resistance and power politics: the experience of non-violent action from Gandhi to the present*. Oxford University Press.

Roberts, A., M. J. Willis, R. McCarthy, and T. G. Ash (2016). *Civil resistance in the Arab Spring: Triumphs and disasters*. Oxford University Press.

Roux, T. (2017). Comparative constitutional studies: Two fields or one? *Annual Review of Law and Social Science 13*, 123–139.

Ryvkin, D. and A. Semykina (2017). An experimental study of democracy breakdown, income and inequality. *Experimental Economics 20*(2), 420–447.

Sausgruber, R. and J.-R. Tyran (2011). Are we taxing ourselves? *Journal of Public Economics 95*(1-2), 164–176.

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press.

Schmidt, K. M. (2009). The role of experiments for the development of economic theories. *Perspektiven der Wirtschaftspolitik 10*, 14–30.

Schotter, A. and B. Sopher (2003). Social learning and coordination conventions in intergenerational games: An experimental study. *Journal of Political Economy 111*(3), 498–529.

Shachar, A. (2009). *The birthright lottery: Citizenship and global inequality*. Harvard University Press.

Sitaraman, G. (2017). *The Crisis of the Middle-Class Constitution. Why Economic Inequality Threatens our Republic*. Knopf.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review 66*(2), 274–279.

Spamann, H. (2015). Empirical comparative law. *Annual Review of Law and Social Science 11*, 131–153.

Stephan, M. J. and E. Chenoweth (2008). Why civil resistance works: The strategic logic of nonviolent conflict. *International security 33*(1), 7–44.

Sutter, M., S. Haigner, and M. Kocher (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *The Review of Economic Studies 77*(4), 1540–1566.

The Economist (January 09, 2016). The arab winter. Available at https://www.economist.com/middle-east-and-africa/2016/01/09/the-arab-winter.

The Norwegian Nobel Committee (October 10, 2015). The nobel peace prize for 2015. Available at https://www.nobelprize.org/prizes/peace/2015/press-release/.

Tobia, K. (2020). Experimental jurisprudence. *Available at SSRN*.

Tyran, J.-R. and L. Feld (2006). Achieving compliance when legal sanctions are non-deterrent. *The Scandinavian Journal of Economics 108*(1), 135–156.

Voigt, S. (2011). Positive constitutional economics II ? A survey of recent developments. *Public Choice 146*(1), 205–257.

Weingast, B. (1997). The political foundations of democracy and the rule of the law. *American Political Science Review 91*(2), 245–263.

Willis, M. J. (2016). Revolt for dignity, tunisia's revolution and civil resistance. In A. Roberts, M. J. Willis, R. McCarthy, and T. G. Ash (Eds.), *Civil resistance in the Arab Spring: Triumphs and disasters*, pp. 30–52. Oxford University Press.

Young, K. (2012). *Constituting economic and social rights*. Oxford University Press.

A Additional Results

Figure A1 shows mean transfers. The level of transfers in transfer (1.99) and transfer-overthrow (2.06) is almost identical.

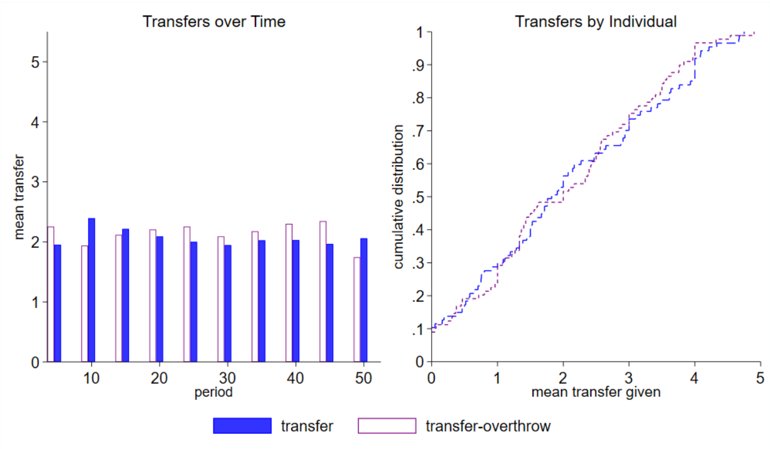


Figure A1: Transfers

Mean transfers over time and cumulative distribution of mean transfers given by each individual.

Figure A2 depicts the means of overall payoffs over time. They do not follow a notable time trend and we can see that the threat of overthrow fails to have a positive impact on overall efficiency.

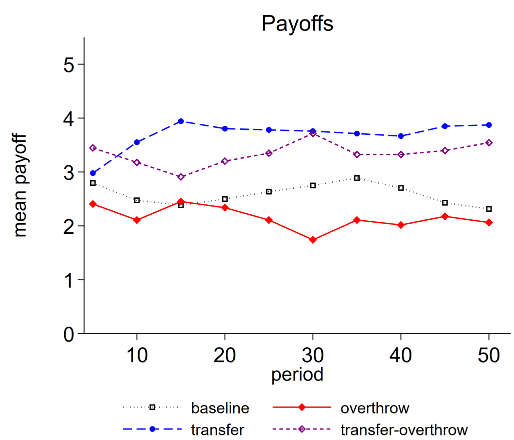


Figure A2: Payoffs over Time

Mean payoffs (after transfers).

Figure A3 shows when an overthrow occurs for the first time. It thus shows how long the initial social order holds up. Almost all groups experience overthrows in the first half of the game.

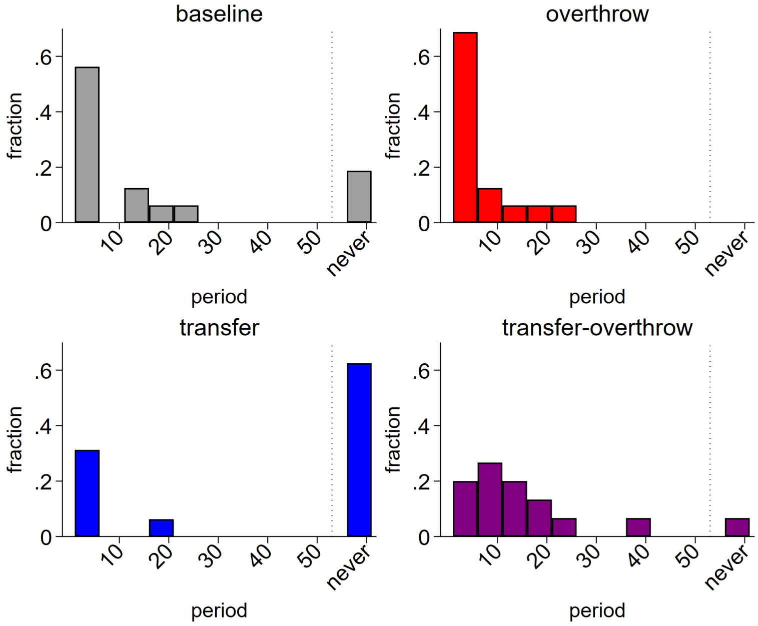


Figure A3: Period of First Overthrow Event

Distribution of the period of the first overthrow event in all groups of a treatment. An overthrow event is a situation in which all matched pairs fail to coordinate on a payoff-yielding equilibrium in a given period.

B Instructions

*Note: The text below shows the instructions for the **baseline** treatment, on which all other treatments build. The additional text in blue was included in the instructions for the **overthrow** as well as the **overthrow-transfer** treatment. The additional text in red was included in the instructions for the **transfer** as well as the **overthrow-transfer** treatment. Instructions displayed here are a translation into English.*[17] *Original instructions where in German and are available from the authors upon request.*

Welcome to our experiment!

If you read the following instructions carefully, you can earn a substantial sum of money, depending on your decisions. It is therefore very important that you read these instructions carefully.

Absolutely no communication with the other participants is allowed during the experiment. Anyone disobeying this rule will be excluded from the experiment and all payments. Should you have any questions, please raise your hand. We will then come to you.

During the experiment, we will speak not of Euros, but of points. Your entire income will therefore initially be calculated in points. The total number of points accumulated by you during the experiment will be paid out to you in Euros at the end, at a rate of:

25 points = 1 Euro.

At the end of the experiment, you will be paid, in cash, the number of points you will have earned during the experiment. In addition to this sum, you will receive payment of 4 Euro for showing up at this experiment.

The experiment consists of at least 50 periods.

After period 50, a draw will decide in each period whether there shall be a further period. With a probability of 75%, there will be a period 51. Should there be a period 51, there will be a period 52 as well, once again with a probability of 75%, etc.

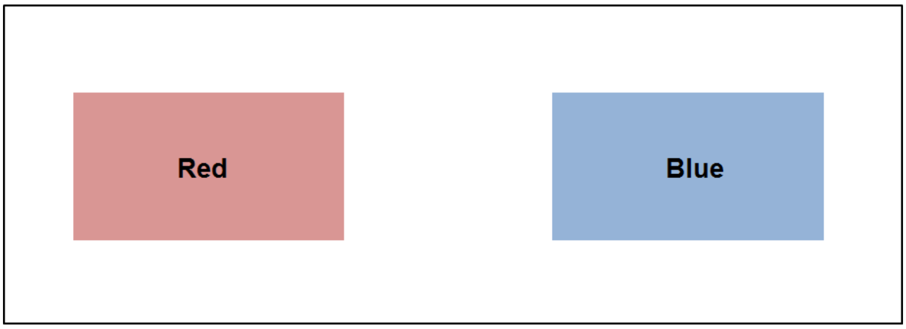
At the beginning of the experiment, participants will be randomly divided into groups of six. Apart from you, your group will therefore be made up of another 5 members. The constellation of your group of six will remain unchanged throughout the entire experiment.

[17] We thank Brian Cooper (MPI Bonn) for the translation.

Also at the beginning of the experiment, the computer will name the participants of each group of six, assigning to each a randomly drawn letter (a, b, c, d, e, or f). Each participant in the group is equally likely to receive a particular letter (a, b, c, d, e, f). Each letter is distributed once in each group of six.

In each period, you will interact with exactly one of the other participants from your group. The computer will randomly determine at the beginning of each period who that other player is. The other 4 participants in your group of six will each also be randomly matched with another participant from the group. In total, there will hence be three parallel encounters in your group in each period.

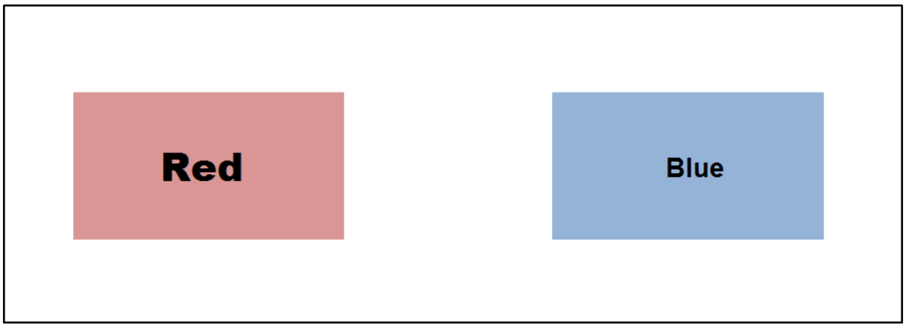
In each period, your task is to choose one of two decision fields:



How many points you earn in a period depends on your decision as well as on the decision of the participant with whom you are interacting.

- If you choose "Red" and the other participant chooses "Blue", you will earn 10 points, and the other participant will earn 1 point.
- If you choose "Blue" and the other participant chooses "Red", you will earn 1 point, and the other participant will earn 10 points.
- If both participants choose "Red", you will both earn 0 points.
- If both participants choose "Blue", you will both earn 0 points.

In each period, one of these fields will be in bold:



Whenever you see the field “Red” in bold, the other participant sees the field “Blue” in bold, and vice versa. You are free to decide whether you wish to follow the marking or not.

The computer decides on the basis of your letter which field is in bold. Whichever participant’s letter comes first in the alphabet sees the field “Red” in bold. If, for example, the computer assigned you the letter c at the beginning of the experiment, and you interact with participant d, e, or f, you will see “Red” in bold. If you interact with participant a or b, however, you will see “Blue” in bold.

For example, if you were assigned the letter a at the start, “Red” will be in bold in all periods. If you are participant f, “Blue” will always be in bold, etc.

Only once both participants have made their decisions will you find out what the other participant has chosen.

At the end of each period, your computer screen will give you an overview of:

- which field you have opted for;
- which field the other participant has chosen;
- the income you and the other participant have each earned in this period;
- how the participants of the other encounters have chosen.

Further, you have the chance to transfer to the other participant any part of your income from the current period. To do this, enter on your screen the number of points you wish to transfer to the other participant, and confirm your entry by clicking “Continue”. You are free to decide whether or not you wish to transfer points and, if you do, how many points you wish to donate.

In addition, your computer screen will give you an overview of:

- the number of points transferred by you to the other participant;

- the number of points the other participant has transferred to you;
- the income you have earned in this period, after transfers;
- the total income you have made up to now;
- the total income each of the other participants in your group of six has made so far

At the end of each period, the letters that were randomly allotted at the beginning of the experiment (a, b, c, d, e, f) can be allotted anew by the computer. As in the beginning, every participant has the same probability of receiving a specific letter (a, b, c, d, e, f). This probability is independent from the letter that he had so far.

This reallotment takes place, if in all three matches within your group of six both participants choose the same color. For instance, if in all three matches both participants choose "Red". Or, for instance, if in two matches both choose "Red" and in the other match both choose "Blue".

In the next period, you will once again be randomly matched with one other participant from your group of six, with whom you will then interact.

Do you have any questions? If yes, please raise your hand. We will come to you.

C Post-Experimental Tests

Other-regarding preferences. Other-regarding preferences were measured using the social value orientation (SVO) ring measure (Liebrand and McClintock 1988). The ring measure consists of 15 modified dictator games in which players allocate money between themselves and another player. Players are characterized on a continuous type space ranging from competitiveness and individualism to prosociality and altruism. To make the *Social Value Orientation (SVO)* comparable to the other scales, we divide the ring degree by 45, such that 0 describes a perfectly selfish (0 degree) individual, and 1 a perfectly pro-social (45 degree) individual.

Risk. The risk elicitation consisted of 1 general risk question and 6 domain-specific questions. The general question read: "Are you, generally speaking, a person willing to take risks or do you rather try to avoid risks?" The domain-specific questions read: "How would you rate your willingness to take risks in the following domains...(i) when driving a car, (ii) in financial investments, (iii) in leisure and sports, (iv) in your carreer, (v) concerning your health, (vi) when trusting unfamiliar people?". There are 10 answer options ranging from *not at all willing to take risks* to *very willing to take risks*. We use the arithmetic mean over the 7 questions as our measure of an individual's risk attitude.

Trust. The trust questions read: "Please rate the following three statements: (i) Generally, people can be trusted. (ii) Nowadays you cannot trust anybody. (iii) When dealing with strangers it's better to be careful before trusting them." The answer options are: *fully agree*, *rather agree*, *rather disagree*, *fully disagree*. The composite trust measure is the arithmetic mean of the three question whereby the first is coded negatively in the other two positively.

Previous Munich Papers in Political Economy:

Betz, Timm and Amy Pond. “Political Ownership”, MPPE No. 1/2020, Munich. (ISSN)2701-3456

Chatziathanasiou, Konstantin, Hippel, Svenja and Michael Kurschilgen. “Property, Redistribution, and the Status Quo. A laboratory study”, MPPE No. 2/2020, Munich. (ISSN)2701-3456

Becker, Annette, Hottenrott, Hanna and Anwesha Mukherjee. “Division of Labor in R&D? Firm Size and Specialization in Corporate Research”, MPPE No. 3/2020, Munich. (ISSN)2701-3456

Steinert, Janina Isabel, Satish, Rucha Vasumati, Stips, Felix and Sebastian Vollmer. “Commitment or Concealment? Impacts and Use of a Portable Saving Device: Evidence from a Field Experiment in Urban India, MPPE No. 4/2020, Munich. (ISSN)2701-3456

Janz, Nicole and Luca Messerschmidt. ”Unravelling the `race to the bottom' argument: How does FDI affect different types of labour rights?”, MPPE No. 5/2020, Munich. (ISSN)2701-3456.

Chowdhury, Subhasish M., Esteve-González, Patricia and Anwesha Mukherjee. ”Heterogeneity, Leveling the Playing Field, and Affirmative Action in Contests”, MPPE No. 6/2020, Munich. (ISSN)2701-3456

Drobner, Christoph. ”Motivated Beliefs and Anticipation of Uncertainty Resolution”, MPPE No. 7/2020, Munich. (ISSN)2701-3456

Chatziathanasiou, Konstantin, Hippel, Svenja and Michael Kurschilgen. “Do rights to resistance discipline the elites? An experiment on the threat of overthrow”, MPPE No. 8/2020, Munich. (ISSN)2701-3456

Impressum:

ISSN: **2701-3456**
Editors: Tim Büthe, Hanna Hottenrott
Associate Editors: Timm Betz, Sebastian Goerg, Eugénia da Conceição Heldt, Michael Kurschilgen, Amy Pond, Sebastian Schwenen, Janina Steinert, Matthias Uhl
Managing Editor: Luca Messerschmidt

Contact: Technicial University of Munich, Arcisstraße 21, 80333 München
mppe@gov.tum.de, mppe@wi.tum.de
https://www.wi.tum.de/mppe/
Twitter: @MunichPapers