Experimental Economics (2021) 24:919–951
https://doi.org/10.1007/s10683-020-09685-5

ORIGINAL PAPER

Property, redistribution, and the status quo: a laboratory study

Konstantin Chatziathanasiou[1] · **Svenja Hippel**[2] · **Michael Kurschilgen**[3,4]

Received: 15 April 2020 / Revised: 30 July 2020 / Accepted: 24 September 2020 /
Published online: 15 October 2020
© The Author(s) 2020

Abstract

We report experimental evidence showing a positive effect of redistribution on economic efficiency via the self-enforcement of property rights, and identify which status groups benefit more and which less. We model an economy in which wealth is produced if players voluntarily comply with the—efficient but inequitable—prevailing social order. We vary exogenously whether redistribution is feasible, and how it is organized. We find that redistribution benefits all status groups as property disputes recede. It is most effective when transfers are not discretionary but instead imposed by some exogenous administration. In the absence of coercive means to enforce property rights, it is the higher status groups, not the lower status groups, who benefit from redistribution being compulsory rather than voluntary.

Keywords Redistribution · Property · Status · Correlated equilibrium · Battle of the sexes · Experiment

JEL Classification C72 · C92 · D74 · H23 · P48

We thank Ernesto Dal Bó, Marco Casari, Christoph Engel, Willemien Kets, Oliver Kirchkamp, Tatiana Kornienko, Ulrike Malmendier, Isabel Marcin, Moti Michaeli, Rosemarie Nagel, J. J. Prescott, Daniel Salicath, Cornelius Schneider, Sebastian Siegloch, and Theresa Wenning, as well as seminar participants at U Bayreuth, U Marburg, U Passau, and LMU Munich for valuable comments on earlier versions of the paper. Financial support of the Max Planck Society is gratefully acknowledged.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10683-020-09685-5) contains supplementary material, which is available to authorized users.

✉ Michael Kurschilgen
m.kurschilgen@tum.de

Extended author information available on the last page of the article

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

1 Introduction

How does redistribution affect a person's economic status? Conceiving redistribution simply as a means to channel wealth from the relatively rich to the relatively poor, the answer is pretty straightforward: it helps the poor, and hurts the rich. Yet the truth is likely to be more complex. A large strand of research in public and monetary economics describes negative behavioral responses to redistribution like lower labor supply, lower effort, and higher expenditures for tax professionals (for an overview of the vast theoretical and empirical literature, see Trabandt and Uhlig (2011), Saez et al. (2012), and Doerrenberg et al. (2017)). If the dead-weight loss is large, redistribution could potentially hurt both the rich and the poor. On the other hand, redistribution might also have a positive effect on economic efficiency by reducing conflict over property rights (Grossman 1994, 1995; Bös and Kolmar 2003; Dal Bó and Dal Bó 2011). But to this day there is no causal empirical evidence for such an effect. The present paper aims to fill that gap. Using a novel experimental paradigm, we test how redistribution affects efficiency via the self-enforcement of property rights, and identify which status groups benefit more and which less.[1]

More effective self-enforcement could free up resources otherwise tied to enforcing property rights by coercive means, for more productive use. Expenditures for deterrence and coercion (police, judiciary, prisons, fences and walls, private security, etc.) are inherently unproductive, and thus socially wasteful (Skaperdas 1992; Hirshleifer 1995). In fact, even in countries with expansive (and expensive) enforcement institutions, property rights are not perfectly secure.[2] The US Department of Justice, for instance, reports for 2018 a property crime rate of 108 victimizations per 1000 households.[3]

Extracting causal evidence from historical or contemporary field data on this important question is extraordinarily difficult. Both redistribution, property rights, and law enforcement are endogenously determined through the political process. In today's market democracies, a person's economic status results from a mix of exogenous factors like inheritance and descent (Bowles and Gintis 2002; Kahlenberg 2010; Chetty et al. 2014; Adermon et al. 2018) and endogenous factors like effort and acquired skill. Moreover, economic status may come along with the power to coerce others and to bend the rules of society to one's advantage (Glaeser et al.

[1] For the purpose of clarity, we distinguish between a person's *status* as the degree of innate privilege in the prevailing social order, and her *economic status* as her position in the distribution of income (Bowles and Gintis 2002).

[2] According to the World Prison Brief of the University of London, the US has an incarceration rate of 655 per 100,000 inhabitants. For comparison, the UK has 148, Germany 77. See https://www.prisonstudies.org/highest-to-lowest/prison_population_rate?field_region_taxonomy_tid=All.

[3] The 108 victimizations include only non-violent property crimes like burglaries, residential trespassing, motor-vehicle thefts, and other thefts. In addition, many violent crimes are also property related. For 2018, the DOJ reports 2.1 instances of violent robbery and 18.4 assaults (excluding rape and sexual assault) per 1000 individuals age 12 or older. See http://www.bjs.gov/index.cfm?ty=pbdetail&iid=6686. Exploiting spatial and temporal variation of land titles in the Brazilian Amazon between 1997 and 2010, Fetzer and Marden (2017) document the effect of insecure property rights on land-related violence, attributing 280 murders directly to land disputes.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

2003; Acemoglu et al. 2015). To generate causal evidence, we therefore design a laboratory environment with (a) no coercive enforcement of property rights, (b) exogenous variation of redistribution, and (c) an exogenous status measure that is orthogonal to other individual characteristics like preferences, productivity, and coercive power.

In particular, we model an *N*-players society of strangers whose members regularly experience anonymous bilateral encounters with one another. Wealth is produced by avoiding disputes over property rights, for which players need to voluntarily agree on who *claims* property of a coveted resource, and who *concedes*. The resulting stage game is a *Battle-of-the-Sexes (BoS)*. An individual's status in society is determined by the *status quo*, reflecting some prevailing legal or social order: a pre-birth lottery 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 status quo is to claim (concede).

We illustrate that, with standard preferences, the status quo functions as a correlation device (Aumann 1974, 1987), and enables frictionless, efficient coordination of otherwise conflicting claims. The correlated equilibrium has a *bourgeois* character (Bhaskar 2000; Gintis 2007), as players comply with the prevailing order and concede to whoever is higher on the ladder. Thus—in equilibrium—the pre-birth status order becomes a self-enforced convention for allocating individual property rights between all members of society; a person's pre-birth status determines her economic status.[4] But in the presence of behavioral types (Embrey et al. 2015) who deviate from the prescribed order, the convention is fragile. The lower a player's rank, the lower her incentives to stick to the order. Theoretically, redistribution stabilizes the bourgeois equilibrium by increasing players' tolerance against occasional deviators.[5]

In a series of experimental treatments, we vary whether redistributive transfers are feasible or not, and compare the effectiveness of different ways to organize redistribution, reflecting stylized transfer institutions with varying degrees of centralization: societies that rely predominantly on alms, tipping, and charity, versus societies with highly centralized welfare states. Specifically, we focus on two dimensions of centralization: (a) whether transfers are paid directly to the beneficiary or indirectly, via a central redistribution pool, and (b) whether players have full discretion over the amount they transfer or transfers of the very same amount are exogenously imposed by some central administration. The comparison

[4] In this theoretical framework, property rights are thus not a constraint but an outcome. For a similar perspective on property rights, see Grossman and Kim (1995); Grossman (2001), and more recently Diermeier et al. (2017).

[5] From a bargaining perspective (Schelling 1956; Crawford 1982), the bourgeois equilibrium can be understood as a self-enforced social contract (Binmore 1998), in which the members of a society have reached a (tacit) agreement about the appropriate compensation for waving one's own claim, in exchange for conceding possession to whoever is higher on the status ladder. Transfers expand the contract zone, i.e. the set of possible agreements. Possession becomes property by mutual acceptance (Bös and Kolmar 2003).

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

between a direct vs. a pooled transfer institution indicates how agents' willingness to provide transfers reacts to diffusion of responsibility (Dana et al. 2007; Hamman et al. 2010; Bartling et al. 2014). Comparing discretionary vs. exogenously imposed transfers yields insights into the mechanism through which transfers affect a player's willingness to concede. As we hold average transfer levels and thus payoff asymmetry constant by design, differences in effectiveness cannot stem from distributional concerns. Instead, they would suggest a different perception of transfers coming from an impersonal process rather than from another human being.

We find that (1) in the absence of redistribution institutions, the status quo translates into an inefficient, pronounced payoff hierarchy. The Gini coefficient is .30 and efficiency reaches only 47% of its potential as players' willingness to concede decreases with their rank on the status ladder. (2) Voluntary redistribution (both direct and indirect) makes all ranks better off as property disputes recede. The Gini coefficient drops to .18 and efficiency increases to 67%. Groups with higher willingness to transfer and thus lower payoff asymmetry systematically achieve higher levels of efficiency. (3) Whereas the effectiveness of voluntary transfers stagnates around 70%, property disputes continue to recede over time when transfers are exogenously imposed, reaching 85% efficiency in late rounds. (4) Virtually all the added surplus of the exogenously imposed redistribution accrues to the upper half of the pre-birth status ladder. (5) By keeping the transfer volume constant, we identify an additional obstacle to the self-enforcement of property rights, besides payoff asymmetry. Exogenously administered, compulsory redistribution is more effective at persuading lower ranks to obey the rules of the status quo by mitigating the dissatisfaction from receiving disappointing transfers.

In sum, our paper documents the existence of a Pareto-improving effect of redistribution via the self-enforcement of property rights. We thus corroborate the conjectures of a strand of theoretical contributions in political economy. Grossman (1994), for instance, models a society of landowners, who earn rents from their land holdings, and peasants, who choose between allocating time to wage employment (on the landlord's premises), self-employment (on their own land), and banditry. He shows that when the technology of banditry is sufficiently effective, redistribution by means of a land reform is the landowners' optimal response to the threat of violent appropriation by the peasants. Grossman (1995) applies a similar argument to identify the conditions under which a class of capitalists voluntarily agrees to redistribute income to the working class via a tax-financed wage subsidy. Acemoglu and Robinson (2000) interpret the extension of voting rights to wider segments of society during the nineteenth century as a strategic commitment to redistribution aimed at preventing a revolution. In Bös and Kolmar (2003), two individuals differ with respect to their initial land possessions, production technology, and appropriation technology. When the time horizon is infinite, a self-enforced, Pareto-improving agreement is possible in which the less productive individual waives his property claims in exchange for a compensatory transfer. More recently, Dal Bó and Dal Bó (2011) illustrate how policies that are distortionary under the assumption of perfectly secure property rights can be optimal in a second-best world (Lipsey and Lancaster 1956) of imperfect property rights, in which such policies "buy social peace".

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Within the production, appropriation and defense framework, several experimental studies have explored the emergence of endogenous property rights. Durham et al. (1998) find evidence on the *paradox of power* (Hirshleifer 1991), i.e. that the initially poorer side will improve its relative position vis-à-vis an initially richer and thus stronger opponent. Carter and Anderton (2001) report strong support for the emergence of stable anarchic equilibria as predicted by the predator-prey model of Grossman and Kim (1996), in which the degree of (socially wasteful) appropriative activity depends on the relative effectiveness of the predator's conflict technology compared to the prey's. In contrast, Duffy and Kim (2005) find that large groups struggle to coordinate on anarchic equilibria as players under-invest in defense leading to excessive predation. The presence of a dictator, who imposes a certain level of defense upon the players, allows groups to coordinate on a Pareto-superior equilibrium. In a cross-cultural lab experiment, Ahn et al. (2016) show that secure property rights are more likely to emerge in countries with high levels of trust, and Ahn et al. (2018) find that communication significantly increases efficiency in a property rights dilemma.

While in the same spirit as those papers, our theoretical framework is more parsimonious, in the interest of keeping the experimental environment sufficiently tractable. First, rather than allowing discrete investments into productive and appropriative activities, our production function relies on a binary action space *(claim, concede)*. Wealth is produced whenever there are secure individual property rights, which occurs when exactly one player claims and the other concedes. Second, our players do not differ in terms of their (production or appropriation) skills but only with respect to their position on the pre-birth status ladder, which defines the relative frequency of being the focal (Schelling 1960) claimant throughout one's life.

From a more technical vantage point, we relate to the literature on coordination of conflicting interests. Similar to the present paper, Isoni et al. (2013) interpret successful coordination on the pure equilibria of a highly asymmetric BoS game as "property conventions". Crawford et al. (2008) show that the power of focality—so effective when players' interests are perfectly aligned (Mehta et al. 1994; Bardsley et al. 2009)—is considerably reduced as soon as payoffs are minimally asymmetric. When payoff asymmetry is very pronounced, i.e. when players differ strongly in their preference ranking over the set of equilibria, even explicit recommendations to play a specific equilibrium fail, leading to substantial efficiency losses (Anbarcı et al. 2018). The recommendations fail because they are largely not followed by the players asked to play their less preferred equilibrium. We show that self-enforced redistribution institutions can restore the power of focality.

Several laboratory experiments have documented that—in repeated 2-person BoS games with partner matching—people use shared common history to successfully coordinate on turn-taking equilibria (Rapoport et al. 1976; Sonsino and Sirota 2003; Arifovic and Ledyard 2018), even if it implies ignoring readily available, exogenous correlation devices (Duffy et al. 2017). In such turn-taking equilibria, the player who concedes the right of playing her preferred equilibrium to her counterpart, relies on (the expectation of) direct reciprocity to (rightfully) expect being compensated by her counterpart's conceding in the future. But as social groups become larger and players interact with varying counterparts, it becomes disproportionately more

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

difficult to (a) construct a shared common history and (b) rely on direct reciprocity. The self-enforced redistribution institutions studied in this paper illustrate how the general ideas of concession and compensation can extend to more anonymous settings.

The growing experimental literature on redistribution has almost exclusively focused on the effects of redistribution when property rights are secure. Agranov and Palfrey (2015) show that when inequality stems from experimentally induced differences in labor productivity, there is an equity-efficiency tradeoff as higher redistribution leads to lower labor supply. Instead, we study a situation without property right enforcement, in which inequality stems from agents being differently privileged in the prevalent order (status quo). Baranski (2016) studies how redistribution affects individual investment decisions into a common project. He shows that adding a second stage where players redistribute the total value of common production via multilateral bargaining yields much higher levels of efficiency than the typical voluntary contributions mechanism (VCM), in which distribution is exogenously imposed. A notable exception is the game studied by Ryvkin and Semykina (2017) where citizens can choose to replace a democratic regime, in which property rights are secure and redistribution requires a majority vote, by an autocrat who promises full redistribution but who can potentially expropriate the citizens. They show that subjects are more likely to voluntarily switch from democracy to autocracy when inequality is high.

The experimental studies of Sausgruber and Tyran (2011), Esarey et al. (2012), and Durante et al. (2014) indicate that redistribution choices in the laboratory are largely in line with observational field data (Fong 2001; Alesina and Angeletos 2005; Alesina and Giuliano 2011). Recent work of Cohn et al. (2019) confirms the explanatory power of lab methods for understanding the redistributive preferences not only of the general population but also of the economic elite. They report that the richest 5% of the US population is less supportive of redistribution than the bottom 95% because a larger share of the top 5% regards unequal earnings as fair even when the inequality is caused purely by luck. Taking advantage of Swiss direct democracy, Epper et al. (2020) document that preferences elicited in the laboratory predict individuals' support for redistribution in several national plebiscites.

Section 2 introduces the theoretical framework, followed by the experimental design in Sect. 3. We present our experimental results in Sect. 4 and discuss our findings in Sect. 5.

2 Theoretical framework

We conceptualize society as a group of N players, whose members regularly experience anonymous bilateral encounters with one another, in which they produce wealth by voluntarily agreeing on who *claims* property of a coveted resource, and who *concedes*. Players differ only with respect to their position on a pre-birth status ladder. We first explain the stage game and then the supergame, in the absence of redistribution opportunities. We characterize a *bourgeois* equilibrium in which lower ranks voluntarily concede to higher ranks, and examine the role of zero-sum transfers in sustaining that equilibrium in the presence of behavioral types.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Fig. 1 Stage game

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

2.1 Stage game

Two members $i = 1, 2$ of the society meet and face the question who of the two should own a coveted resource. For each player i, the set of possible actions is $A_i = \{claim, concede\}$. We define $a = (a_1, a_2)$ as an action profile with a_1 being the action of player 1 (the row player) and a_2 being the action of player 2 (the column player). Each player prefers being the sole claimant (*claim*, *concede*) to the other player being the sole claimant (*concede*, *claim*) to having unproductive disagreement (*claim*, *claim*) or (*concede*, *concede*). Figure 1 shows the normal form game G with payoffs $x_i \in \{0, l, h\}$ and $0 < l < h$.[6]

The two pure Nash equilibria of G are $e_1 = (claim, concede)$ and $e_2 = (concede, claim)$, where e_1 is the equilibrium more favorable to player 1, and e_2 to player 2.[7] The mixed equilibrium e_{mix} is constituted by playing the action *claim* with $P_{mix}(claim) = \frac{h}{h+l}$ and results in an expected payoff of $E_{mix} = \frac{hl}{h+l} < l$.[8] The mixed equilibrium is thus unsatisfactory, both from a social and from an individual perspective whereas the two pure equilibria pose a coordination problem in which each of the players will eye her preferred equilibrium.

Potentially, this situation could be resolved with the help of a correlation device (Aumann 1974, 1987). This could for instance be a recommendation of play derived from some prevailing legal or social order, i.e. the *status quo*. Whenever two players meet, they know who is supposed to claim, and who to concede. For instance, the prevailing order could stipulate who is supposed to claim a certain piece of land (the first-born son of the deceased former owner or the peasant who has worked that land for years). It could also stipulate who is supposed to claim medical treatment at a crowded hospital (the person with the more expensive health plan or the person with the more urgent medical condition).

Definition 1 $\phi = (M_1, M_2, \pi)$ is a direct correlation device, where M_i, $i = 1, 2$, is the finite set of messages $M_i = A_i$ for player i. There is a probability distribution

[6] In some real-world examples, (*concede*, *concede*) may be preferable to (*claim*, *claim*). Increasing the payoff of (*concede*, *concede*) to l instead of 0 does not affect the general structure of the game, see "Appendix A.2".

[7] Note that in this conception of the BoS, players coordinate by choosing *different* actions.

[8] 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 $P(concede, concede) = \frac{l^2}{(h+l)^2}$. Disagreement is defined as playing either (*claim*, *claim*) or (*concede*, *concede*). The probability of disagreement is thus $\frac{h^2+l^2}{(h+l)^2}$.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

over the set of possible message profiles $M = \{e_1, e_2\}$ with $Prob(e_1) = \pi$ and $Prob(e_2) = 1 - \pi$. The device selects a message profile $m \in M$ according to the probability distribution and privately sends m_i to player i.[9]

In the extended game G_ϕ, where players receive a message before they play G, the combination of an identical dyadic action space A_i for both players and a direct correlation device (Myerson 1994) with mutually exclusive messages leads to the full revelation of the other player's message given one's own private message. Thus, when a player receives one of the two possible messages (*claim* or *concede*), she knows that her counterpart received the opposite message. As a result, mutually following the device will always end up in one of the pure Nash equilibria.

Definition 2 A bourgeois equilibrium is a pair (ϕ, σ) such that the pure strategies $\sigma_i : A_i \to A_i$ of the players are identity maps that constitute a Nash equilibrium of the extended game G_ϕ.[10]

In this correlated equilibrium, players always comply with whatever is recommended in the status quo: $\sigma_i(a_i) = a_i$. Following Bhaskar (2000) and Gintis (2007), we refer to it as the *bourgeois equilibrium*.[11]

2.2 Supergame

The supergame consists of an indefinite series of random and anonymous two-person encounters within the society of N players. In every encounter, the extended game G_ϕ is played. A pre-birth lottery determines a player's rank in society. The higher (lower) one's rank, the more often the action recommended by the status quo is to *claim* (*concede*). A player's rank can be understood as a bundle of pre-birth characteristics (gender, ethnicity, inherited wealth, parents' education, etc.) affecting the frequency of situations in which—in the prevailing social or legal order—an individual is supposed to get a coveted resource, or to accept that somebody else gets it.

Definition 3 Let Θ be an exogenous status hierarchy were $\theta_i \in \{1, 2, \ldots, N\}$ denotes the *rank* of player i in the society such that the lower the number the higher the rank, and no two players can have the exact same rank $\theta_i \neq \theta_{-i}$. The function $\vartheta : \Theta \to \pi$ induces a message-hierarchy. ϑ chooses the probability π of the direct correlation device ϕ such that a player's probability to receive her favorable message increases linearly with her rank, i.e. $\vartheta(\theta_i) = \pi = \frac{N-\theta_i}{N-1}$.

[9] Both the definition of the extended game and of the correlated equilibrium closely follow the notation of Duffy et al. (2017).

[10] A bourgeois equilibrium corresponds to a direct correlated equilibrium following (Aumann 1974, 1987).

[11] In an infinitely repeated, symmetric, 2-person BoS game, Bhaskar (2000) distinguishes between an *egalitarian* convention, in which players use successful (lucky) coordination in the initial period to tacitly agree on alternating between both stage-game equilibria in all subsequent periods, and a *bourgeois* convention, in which players tacitly agree on sticking with the initial stage-game equilibrium forever.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

The rank of a player is not *directly* payoff-relevant. But it becomes *indirectly* payoff relevant through (1) the correlation device favoring higher ranks and (2) the common expectation of compliance with the status quo. The former is due to ϑ determining the probability π of the direct correlation device ϕ such that it sends the message $m_i = claim$ more often to player i, the higher i's rank. The latter, as we have shown above, is true in the bourgeois equilibrium. Thus—in equilibrium—the function ϑ manifests the hierarchical ordering of Θ into a hierarchy of expected payoffs E_{θ_i}:

$$E_{\theta_i} = h\pi + l(1-\pi) = \frac{h(N-\theta_i) + l(\theta_i - 1)}{N-1} \tag{1}$$

In the bourgeois equilibrium, the highest rank $\theta_i = 1$ earns $x_i = h$ and the lowest rank $\theta_i = N$ earns $x_i = l$. The other ranks' expected payoffs fall between these two extremes, strictly (and linearly) decreasing in the rank's number.

Hypothesis 1 *The exogenous status hierarchy translates into a payoff hierarchy.*

2.3 Deviations from the bourgeois equilibrium

Potentially, there are many reasons why players would not follow the recommendation. For instance, there could be strategic uncertainty about the counterparts' level of rationality.[12] Players may just disregard the device or refuse to follow the recommendation due to some non-standard preferences (DellaVigna 2009), for instance, other-regardingness.[13] There could also be misunderstandings about the direct nature of the device, resulting in ambiguity about the interpretation of the message (Duffy et al. 2017). Moreover, beliefs could be strategically distorted (Di Tella et al. 2015). We subsume all of the above reasons into the potential existence of *behavioral types* (Embrey et al. 2015) who sometimes deviate from the bourgeois equilibrium, and define w^j as the probability with which player i expects her counterpart j to deviate in a given encounter.[14] In the spirit of trembling hand perfection (Selten 1975), we then compute tolerance thresholds $\overline{w}^j$ to determine the maximum deviation propensity that a player would tolerate before starting to deviate herself from the bourgeois equilibrium.

In the supergame, a player is willing to comply and receive E_{θ_i} in $(1 - w^j)$ of her encounters and zero otherwise as long as complying with the status quo is more

[12] For example Rosenthal (1989) assumes in his bounded-rationality approach that best replies need only be played with a larger probability than other strategies, but not necessarily with probability 1.

[13] Note that (1) aversion to advantageous inequality (triggered by earning h and the other player l) decreases the distance in utility space between monetary payoffs h and 0. (2) Aversion to disadvantageous inequality (when earning l and the other player h) decreases the distance between l and 0. (3) People dislike disadvantageous inequality more than advantageous inequality. As a result, for inequality averse players, the difference between h and l is even larger in utility space than in monetary payoff space. See, for instance, Fehr and Schmidt (1999) and Charness and Rabin (2002).

[14] Note that for the belief w^j it is immaterial whether a given individual deviates deterministically (refusing to ever play the bourgeois equilibrium) or probabilistically (randomizing over the action space). Rather, w^j captures the mean disposition to deviate in the population.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

profitable than ignoring the device and receiving E_{mix}.[15] The lower a player's rank θ_i, the less frequently (π) she is supposed to *claim* in the status quo, the less she has to lose from the collapse of the bourgeois equilibrium, the lower her tolerance threshold $\overline{w}^j_{\theta_i}$:

$$\overline{w}^j_{\theta_i} = \frac{E_{\theta_i} - E_{mix}}{E_{\theta_i}} = 1 - \frac{hl}{h+l}[\pi h + (1-\pi)l]^{-1} \tag{2}$$

The lowest-ranked player's ($\pi = 0$) threshold $\overline{w}^j_N = \frac{l}{h+l}$ turns out to be the *critical threshold* for the existence of the bourgeois equilibrium. If the lowest rank N decides that it pays more to disobey the status quo, this would trigger a chain reaction that reduces the tolerance thresholds of higher-ranked players likewise to $\overline{w}^j_N$, leading to the collapse of the bourgeois equilibrium. To see this, assume for a moment that the lowest-ranked player systematically disobeyed and always played *claim*. As a consequence, the second-lowest player $N-1$ would see her payoff from the only encounter in which she is higher ranked being reduced from h to 0 (if she continues playing *claim*) or l (if she disobeys the device herself and plays *concede*). Her expected payoff would thus be reduced (at least) to E_N and her tolerance threshold would drop to:

$$\overline{w}^j_{N-1} = \frac{l}{h+l} = \overline{w}^j_N := \overline{w}^{crit} \tag{3}$$

Proposition 1 *If the common belief w^j about the mean disposition to disobey is below $\overline{w}^{crit} = \frac{l}{h+l}$, then there exists a bourgeois equilibrium, in which all players comply with the status quo.*

2.4 Redistribution

The peril of entering such ruinous dynamic could be mitigated by increasing $\overline{w}^{crit}$. In principle, a simple transfer institution could achieve this. Consider for instance the possibility of making a zero-sum transfer in the immediate aftermath of successful coordination in the stage game.

Definition 4 $\mathcal{T}$ is a transfer stage in which players of G_ϕ with payoff $x_i = h$ can make direct transfers $\tau_i \in [0, h]$ to the other player of G_ϕ with payoff $x_j = l$ after successful coordination.

A transfer τ in every encounter would flatten the hierarchy of tolerance thresholds against potential deviations in (2). High ranks' thresholds would go down by the same amount that low ranks' thresholds go up. As a result, the *critical* threshold would increase by $\frac{\tau}{h+l}$:

[15] $E(a_i = m_i) \geq E(P_{mix}(claim)) \Rightarrow w^j 0 + (1 - w^j)E_{\theta_i} \geq E_{mix}$. For a detailed analysis of deviations in the stage game, see "Appendix A.1".

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

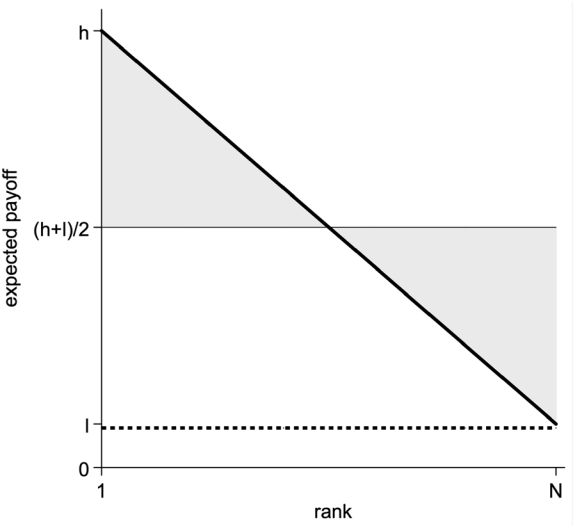


Fig. 2 Equilibrium payoffs by rank. The dashed horizontal line slightly below l denotes expected payoffs in the mixed equilibrium. The solid diagonal line shows expected payoffs in the bourgeois equilibrium without transfers, ranging from an expected payoff of h for rank 1 to an expected payoff of l for rank N. The shaded area up to the solid horizontal line at $\frac{h+l}{2}$ shows the bourgeois equilibrium with varying volume of transfers

$$\overline{w}_{\tau}^{crit} = \frac{l+\tau}{(l+\tau)+(h-\tau)} = \frac{l+\tau}{h+l} \tag{4}$$

In order to persuade the lowest rank to obey the status quo, high ranks need to transfer $\tau_{min} \geq w^j(h+l) - l$. Common knowledge about the mean disposition to disobey w_j thus translates into common knowledge about τ_{min}. The higher w^j, the higher the transfer needed to stabilize the bourgeois equilibrium. If provided, transfers would broaden the set of possible final payoff distributions as shown in Fig. 2. The diagonal line represents the expected payoff distribution in the bourgeois equilibrium with zero transfers. The horizontal line depicts the most extreme form of payoff redistribution (*egalitarian optimum*).[16] The shaded area between these two lines is the set of possible bourgeois equilibria reachable with different levels of average transfers. All ranks are better of in any bourgeois equilibrium than in the mixed equilibrium. The higher the rank, the larger the difference.

Hypothesis 2 *Transfers increase the incidence of the bourgeois equilibrium.*

It can easily be shown that the voluntary provision of τ_{min} can be sustained in equilibrium. If all other players (are commonly expected to) provide exactly τ_{min}, no player has an incentive to unilaterally provide neither more nor less than τ_{min}. A

[16] The highest reasonable transfer is $\tau_{max} = \frac{h-l}{2}$, which would result in all players having equal payoffs. Transfers beyond τ_{max} would create new inequality by reversing the rank hierarchy. τ_{max} allows to stabilize the bourgeois equilibrium when the mean disposition to disobey is $\overline{w}_{max}^{crit} = \frac{1}{2}$. If w^j were even larger than $\overline{w}_{max}^{crit}$, transfers would not be able to stabilize the equilibrium.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

higher transfer would entail additional cost without added benefit since τ_{min} is sufficient to secure perfect obedience of the lowest rank. On the other hand, a *lower* transfer would fail to reach the critical threshold, triggering the collapse of the bourgeois equilibrium. As shown above, all ranks are worse off in the mixed equilibrium. Thus, no player has a reason to deviate from τ_{min}.

Proposition 2 *If the common belief w^j about the mean disposition to disobey is below $\overline{w}^{crit} = \frac{1}{2}$, there exists a bourgeois equilibrium, in which the transfer τ_{min} is provided and all players comply with the status quo.*

3 Experimental design

We conduct a laboratory experiment to test how redistribution affects efficiency via the self-enforcement of property rights, and to identify which status groups benefit more and which less. As laid out in the preceding section, our experimental environment describes an economy with zero coercive means to protect property claims. Status is being exogenously determined in a pre-game lottery, which defines a player's probability of receiving the recommendation to play *claim* or *concede* in a given period.

We compare a baseline treatment (*no-T*), in which redistributive transfers are not possible, to three treatments with an additional transfer stage. In treatment *T-direct* players can voluntarily make direct transfers to the person they just interacted with. In *T-pool* they can voluntarily transfer money to a central pool which is spread equally to all low earners of that period. In *T-admin* players do not have discretion over their transfers. Instead, transfers are exogenously determined by a random draw from the empirical distribution of *T-direct*.

The three transfer treatments reflect stylized transfer institutions with varying degrees of centralization, along two dimensions: (a) whether transfers are paid directly to the beneficiary or indirectly, via a central redistribution pool (*T-direct* vs. *T-pool*), and (b) whether players have full discretion over the amount they transfer or transfers of the very same amount are exogenously imposed by some central administration (*T-direct* vs. *T-admin*).[17] In the following, we explain the treatments in detail, followed by general experimental procedures and behavioral predictions on expected treatment differences.

3.1 Treatments

no-T. The *no-T* treatment implements the game G_ϕ described in Sect. 2. Subjects are randomly assigned to groups of $N = 6$. The composition of the groups is constant

[17] In reality, those institutional differences are generally not *ceteris paribus*. Rather, transfer institutions that vary in terms of centralization, tend to also differ in other aspects, most notably, their degree of coercion, and of administrative efficiency. We deliberately keep those aspects constant.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

over all periods. The paper instructions inform participants that, at the beginning of the experiment, each subject of a given matching group is randomly allocated a unique rank, labeled $\{a, b, c, d, e, f\}$ from highest to lowest. After the random draw, subjects are informed on their computer screens about their own rank. In every period, subjects are randomly paired within the group and play the 2-player BoS-game of Fig. 1 with $h = 10$ and $l = 1$.[18] In the experimental instructions, the two actions *claim* and *concede* are labeled as *red* and *blue*, respectively:

- *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.*

The correlation device is introduced through the salient highlighting of one of the two possible color-coded actions, i.e., the relatively higher-ranked (lower-ranked) player of each encounter sees the action *red* (*blue*) highlighted in bold letters. This is commonly known to all subjects:

> *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). [...] 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.*[19]

Subjects are not explicitly informed about the absolute rank of their counterpart but can always infer the relative rank from the correlation device. There were no persistent individual identifiers shown to the players. At the end of each period, subjects are informed about the choice of their counterpart and the resulting payoffs. In addition, subjects receive information about how the participants of the other encounters (within the same group) have chosen, and about the total cumulated payoff of all participants (within the same group). The actions and payoffs of the other group members were not indexed by their rank letter (nor any other individual identifier).

[18] With these parameters, the highest rank (i.e. the richest 17th percentile) earns 10 times more than the lowest rank (i.e. the poorest 17th percentile) in the bourgeois equilibrium before transfers. This hierarchy of expected payoffs approximates roughly the current income disparity in Germany, where the experiment was conducted. In Germany, the richest 17 percent of the population earn about 12 times more than the poorest 17 percent. For comparison, in the United States the richest 17 percent earn about 50 times more than the poorest 17 percent. Pre-tax income figures from the World Inequality Database: https://wid.world/.

[19] See "Appendix A.3".

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

T-direct. The *T-direct* treatment is identical to *no-T* except that there is now an additional transfer stage (in every period), in which players can simultaneously make direct zero-sum transfers to each other. If a player earned $x_i = 10$ ($x_i = 1$) in G_ϕ, she can transfer up to 10 (1) tokens to the player who earned $x_i = 1$ ($x_i = 10$) in G_ϕ of the same encounter.[20]

T-pool. The *T-pool* treatment is identical to *no-T* except that there is now an additional transfer stage (in every period), in which players can simultaneously make payoff transfers to a fund that is spread evenly among all qualifying players. If a player earned $x_i = 10$ ($x_i = 1$) in G_ϕ, she can transfer up to 10 (1) tokens to pool P_L (P_H). Within each group, the sum of transfers to P_L (P_H) is then distributed equally among all players who earned $x_i = 1$ ($x_i = 10$) in that period.

T-admin. In *T-admin* players cannot choose themselves the amount transferred to their counterpart. Instead, players are informed that the computer will automatically determine a transfer amount. In particular, in each stage game of *T-admin*, the transfer amount τ_i is determined by a random draw from the empirical distribution of transfers in *T-direct*. For that purpose, all transfer decisions (including the decisions to transfer zero) from *T-direct* are put into one of 12 urns Υ_x^θ, depending on the rank $\theta \in \{a, b, c, d, e, f\}$ of the transferring player, and her stage-game payoffs $x \in \{1, 10\}$.[21] Whenever players in *T-admin* succesfully coordinate on (*red*, *blue*) or on (*blue*, *red*) and thus earn payoffs (10, 1) or (1, 10), the experimental software randomly draws (with replacement) a transfer amount from the respective urn. If for instance, *b* succesfully coordinated with *c* on (*red*, *blue*), the software would draw an amount τ_i from urn Υ_{10}^b to be transferred from *b* to *c*, and an amount τ_i from urn Υ_1^c to be transferred from *c* to *b*.[22] As a result, transfers in *T-admin* are virtually identical to transfers in *T-direct*.[23] See Figure A5.

3.2 Procedure

Every subject participated in exactly one supergame, which lasted at least 50 periods. To mimic an indefinite time horizon, we use a random continuation rule (Roth and Murnighan 1978), closely following the protocol of Camera and Casari (2009)

[20] Note that, in addition to transfers from $x_i = 10$ to $x_i = 1$ players, our experimental design also allowed, in principle, transfers from $x_i = 1$ players to $x_i = 10$ players (see "Appendix A.3"). We chose not to preclude the latter in order to keep the transfer stage as normatively neutral as possible. Empirically however, those $x_i = 1$ to $x_i = 10$ transfers are negligible as they account for only $84/3417 = 2.5\%$ and $80/2471 = 3.2\%$ of all tokens transferred in *T-direct* and *T-pool*, respectively.

[21] When payoffs are zero, transfers are not possible.

[22] Remember that the experimental design of *T-direct* also allowed transfers from $x_i = 1$ players to $x_i = 10$ players. Despite that case being empirically irrelevant (virtually all entries in Υ_1^θ are zero), we account for it to keep the experimental instructions of *T-admin* and *T-direct* as similar as possible.

[23] We do not further differentiate the urns by period and/or by the rank of the transfer recipient since those two factors are empirically irrelevant (and have no implications for the wording of the instructions). Moreover, note that we do not inform participants of *T-admin* about the exact procedure that generates the transfers since that information could potentially convey a social norm.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

and Bigoni et al. (2019). Subjects play 50 periods with certainty, thereafter the continuation probability decreases to $\delta = 0.75$. Hence, in every period $t > 50$, the expected number of additional periods is 3. The exact duration of the supergame was randomly determined for each experimental session. The longest session ended randomly after 60 periods.

All periods played were payoff-relevant, i.e. participants were paid the sum of the earnings of all periods played.[24] Before the start of the experiment, paper instructions (see "Appendix A.3") were handed out and read aloud to ensure common knowledge. Additionally, subjects had to pass extensive control questions to ensure full understanding. After the supergame, we elicited participants' (a) other-regarding preferences, (b) risk and trust attitudes as commonly elicited in the German Socio-economic Panel (SOEP), (c) some socio-demographics (age, gender, number of siblings). See "Appendix A.4".

The experiment was conducted at the BonnEconLab of the University of Bonn, Germany, and was computerized using the software z-Tree (Fischbacher 2007). From a database of more than 5000 people, we recruited 384 subjects (96 per treatment), using hroot (Bock et al. 2014). Each subject participated only in one treatment (between-subject design). Subjects were mainly undergraduate students from a variety of disciplines. Sessions lasted about 90 min and subjects earned on average 18.01 € (about 22.00 $) 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.

3.3 Behavioral predictions

The goal of the experiment is to test how redistribution affects efficiency via the self-enforcement of property rights, and to compare the effectiveness of different ways to organize redistribution, varying in their degree of centralization. In the following, we discuss how the different institutional settings are likely to perform depending on the relative strength of several plausible behavioral mechanisms.

As detailed in Sect. 2, if there is reluctance to follow the correlation device because of the resulting payoff asymmetry (10, 1), we should observe substantial miscoordination in the baseline *noT*. Transfer institutions enable the high earners to reduce inequality and thus increase low earners' willingness to coordinate on the bourgeois equilibrium (i.e. choose red when red is shown in bold, and choose blue when blue is shown in bold). In *T-direct*, the earner of 10 can make a transfer to the earner of 1 after successful coordination. Even though the high earner benefits from increasing the probability with which the low earner will concede in a future encounter, she has an incentive to free ride on the transfers of other high earners. Given the lack of persistent individual identifiers, transfers cannot be used to spark

[24] Consequently, participants could earn more in longer sessions than in shorter sessions but (a) they also had higher opportunity costs of staying longer in the lab and (b) the difference between the shortest and the longest session was only about 5 min.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

and sustain cycles of *direct* reciprocity (Gintis 2000; Fehr and Gächter 2000; Fehr et al. 2002).

We compare the effectiveness of *T-direct*, in which transfers are paid directly to the beneficiary, to *T-pool*, in which transfers are paid indirectly, via a central redistribution pool. Within the rational choice framework of Sect. 2, the transfers schemes of *T-direct* and *T-pool* do not differ in their incentive structure. Both pose the exact same problem of collective action. Behaviorally, however, they could differ. In fact, whether subjects are more willing to provide transfers in *T-direct* or in *T-pool* might depend on whether high earners conceptualize the decision to transfer from a *backward-looking* or from a *forward-looking* perspective.

From a backward-looking perspective, a transfer would be given to reward the behavior of one's recent counterpart; i.e. to *return* either trust or the favor of a gift exchange (Berg et al. 1995; Fehr et al. 1998; McCabe et al. 1996; Gächter and Falk 2002; Engelmann and Ortmann 2002). In this case, the *T-direct* mechanism should trigger higher willingness to transfer since it allocates the entire amount to the person targeted whereas the mechanism of *T-pool* dilutes the expression of gratitude, and diffuses the responsibility (Dana et al. 2007; Hamman et al. 2010; Bartling et al. 2014) for rewarding the obedient lower rank. From a forward-looking perspective, a transfer would be given as an investment into the future obedience of a likely interaction partner; i.e. as an attempt to *initiate* a cycle of *indirect* reciprocity (Alexander 1987; Boyd and Richerson 1989; Dufwenberg et al. 2001; Greiner and Levati 2005; Nowak and Sigmund 2005). In this case, the *T-pool* mechanism is more appealing since it hedges the investment risk between several possible future counterparts.

In sum, we should observe higher average transfers and better coordination in *T-direct* (*T-pool*) if the backward-looking (forward-looking) perspective dominates. In addition, even if average transfers in *T-direct* and *T-pool* were identical, some high earners are likely to be more generous than others. The *T-pool* mechanism smooths out the period-by-period variation of the transfers received by individual low-ranked players, making it easier for players to learn the expected average transfer, potentially enabling them to reach the equilibrium faster.[25]

Finally, we compare a setting in which players have full discretion over the amount they transfer (*T-direct*) to one in which transfers of the very same amount are exogenously imposed by some central administration (*T-admin*). If only distributional concerns mattered, we should see no difference between both treatments as average transfer levels are identical by design. However, transfers between *T-admin* and *T-direct* could differ if indirect reciprocity matters. In *T-admin*, players cannot reciprocate to receiving transfers since the transfers are exogenous (Charness and Rabin 2002; Falk et al. 2008). Whether transfers in *T-direct* trigger *positive* or rather *negative* indirect reciprocity will depend on the low earner's perception of the transfer received *relative* to what she deems appropriate. If the average transfer level is perceived positively (i.e. more than appropriate), low ranks will express their gratitude by conceding more in *T-direct* than in *T-admin*. On the other hand, if the average transfer level is perceived negatively (i.e. less than appropriate), low ranks

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

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

will express their dissatisfaction by conceding less in *T-direct* than in *T-admin*. In addition, players in *T-direct* could be tempted to *strategically* withhold conceding (Schelling 1956; Crawford 1982), particularly early in the game, in an attempt to press the high ranks for more generous transfers. Strategic considerations are pointless in *T-admin*.

4 Results

We first report how redistribution affects overall efficiency and the economic status of different ranks. Subsequently, we look into subjects' willingness to concede to higher-ranked players and their willingness to transfer money to lower-ranked players as the behavioral drivers of treatment differences.

4.1 Payoffs

Figure 3 shows the development of efficiency over time, in payoff units. While the *no-T* treatment generates substantially more wealth than the mixed equilibrium, it achieves only 47% of the payoffs attainable in the bourgeois equilibrium. With redistribution, efficiency improves to 67% in *T-direct* ($p = .003$) and 64% in *T-pool* ($p = .016$) but still falls significantly short of the bourgeois equilibrium prediction. In the first 15 periods, all three transfer treatments display a noticeable increase in efficiency whereas the *no-T* treatment does not. From period 15 on, payoffs in *T-direct* and *T-pool* stagnate at or below 70%. In contrast, coordination in *T-admin* continues to rise throughout the duration of the supergame, averaging 85% in the last 10 periods, significantly higher than *T-direct* ($p = .041$) despite having—by design—the same level of transfers.[26]

All these observed treatment differences emerge over time. In period 1, there are no significant differences between *no-T* and any of the transfer treatments ($p > .182$). Neither is there a difference between *T-direct* and *T-admin* ($p = .183$). Interestingly, initial coordination (and thus efficiency) is weakly higher in *T-pool* than in *T-direct* ($p = .086$). But this difference vanishes over time.

[26] Unless otherwise specified, all our statistical tests are guided by the following principles: (1) Since the random continuation rule varied the duration of each experimental session, we only report results from the first 50 (guaranteed) periods of the game. (2) In period 1, each individual is an independent observation. From period 2 on, each group is an independent observation. Per treatment, there are 96 subjects, i.e. 16 statistically independent groups. (3) For comparisons between treatments we report P-values of two-sided Mann-Whitney rank-sum tests over group means. (4) For comparisons of observed behavior with theoretical predictions, we report P-values of two-sided Wilcoxon signed-rank tests over group means. (5) For comparisons between different ranks of the same treatment, we report P-values of two-sided Wilcoxon signed-rank tests of matched pairs. Decisions are first averaged over 50 periods, by individual. The means of rank i and rank j are then matched by group. For each comparison of any two ranks, there are thus 16 matched-pairs per treatment.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Result 1 *Transfer institutions lead to better coordination and thus higher efficiency. Exogenously administered transfers are the most effective.*

Figure 4 shows payoffs over time separately for the lower ranks ($d - f$) and the upper ranks ($a - c$). Panel A shows that for the lower half of the status hierarchy, the exact design of the transfer institution is immaterial. The large payoff difference concerns the presence or absence of some (any) transfer institution. In contrast, for the upper half the exact design matters a lot. Players who were given high status in the pre-play lottery benefit considerably from not having discretion over the amount they transfer. The difference between *T-admin* and *T-direct* begins to emerge around period 15 and continues to grow toward the end of the game. In the last 10 periods, upper ranks earn 26 percent more in *T-admin* than in *T-direct* ($p = .026$).[27]

Result 2 *Upper ranks benefit from not having discretion over the amount they transfer. For the lower ranks, the exact design of the transfer institution is immaterial.*

In Fig. 5 we disaggregate the treatment effects even further, to identify individual ranks and compare payoffs *before* (left) and *after* (right) transfers. The dashed black line at .9 shows the predicted payoffs in the mixed equilibrium: such a society would be poor but egalitarian. In contrast, the average group of the *no-T* treatment is significantly richer ($p < .001$) and displays a pronounced payoff hierarchy. The higher a player's rank, the higher her payoff.[28] Moreover, no rank is worse off than in the mixed equilibrium. Each rank $a - d$ earns significantly more than the mixed equilibrium ($p < .049$) payoff.

There are four interesting observations. First, the availability of transfer institutions flattens the payoff hierarchy. The Gini coefficient drops from .30 in *no-T* to about .18 in the transfer treatments ($p < .002$).[29] But even after transfers, the distribution of payoffs is far from flat but still conditioned by the draw of luck that determines a player's position on the status ladder. Second, lower ranks not only benefit *directly* from the institutional environment—through the net transfers received—but also *indirectly*—through the reduction of miscoordination. In fact, compared to the *no-T* treatment, in each of the transfer treatments net recipients d ($p < .006$) and e ($p < .023$) would already be significantly better off *before* transfers, i.e. solely through the reduction of disputes over claims. Rank f, however, is only better off *after* transfers ($p < .001$). Third, no rank (not even a) is better off without rather than with transfer institutions. Averaged over all periods, even net contributors b ($p < .055$) and c ($p < .003$) have substantially higher after-transfer payoffs in *T-direct* and *T-admin* than in *no-T*. Fourth, no rank loses from transfers

[27] Note that when averaging over *all ranks* (as in Fig. 3), 5.5 is the highest possible mean payoff. *Subgroups* of ranks can achieve mean payoffs above 5.5 (as in Panel B of Fig. 4).

[28] Averaged over all 50 periods, in the *no-T* treatment a earns substantially more than f ($p < .001$); among the directly adjacent ranks, a earns more than b ($p = .020$), b more than c ($p = .017$), c more than d ($p = .038$), d more than e ($p = .002$), e more than f ($p = .083$)

[29] For comparison, the bourgeois equilibrium (without transfers) and the mixed equilibrium generate Gini coefficients of .32 and 0 respectively.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

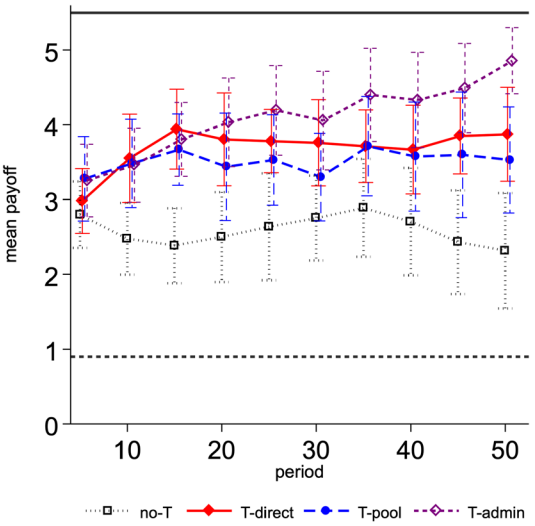


Fig. 3 Payoffs over time. Mean payoffs after transfers. The solid (dashed) horizontal line at 5.5 (.9) denotes predicted average payoffs in the bourgeois equilibrium (mixed equilibrium). The 95% confidence intervals capture the between-group variation around the treatment means

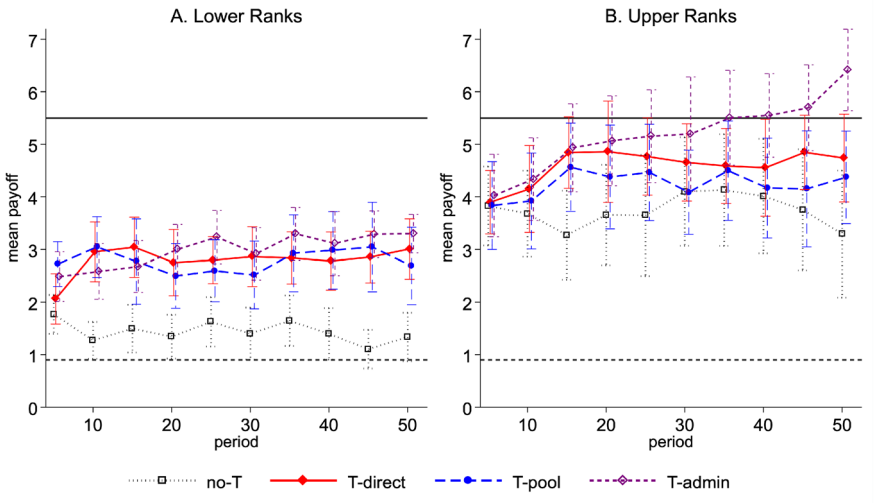


Fig. 4 Payoffs over time—lower vs. upper ranks. Mean payoffs after transfers, for **A** ranks $d-f$, **B** ranks $a-c$. The solid (dashed) horizontal line at 5.5 (.9) denotes predicted average payoffs in the bourgeois equilibrium with maximum redistribution (mixed equilibrium). The 95% confidence intervals capture the between-group variation around the treatment means

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

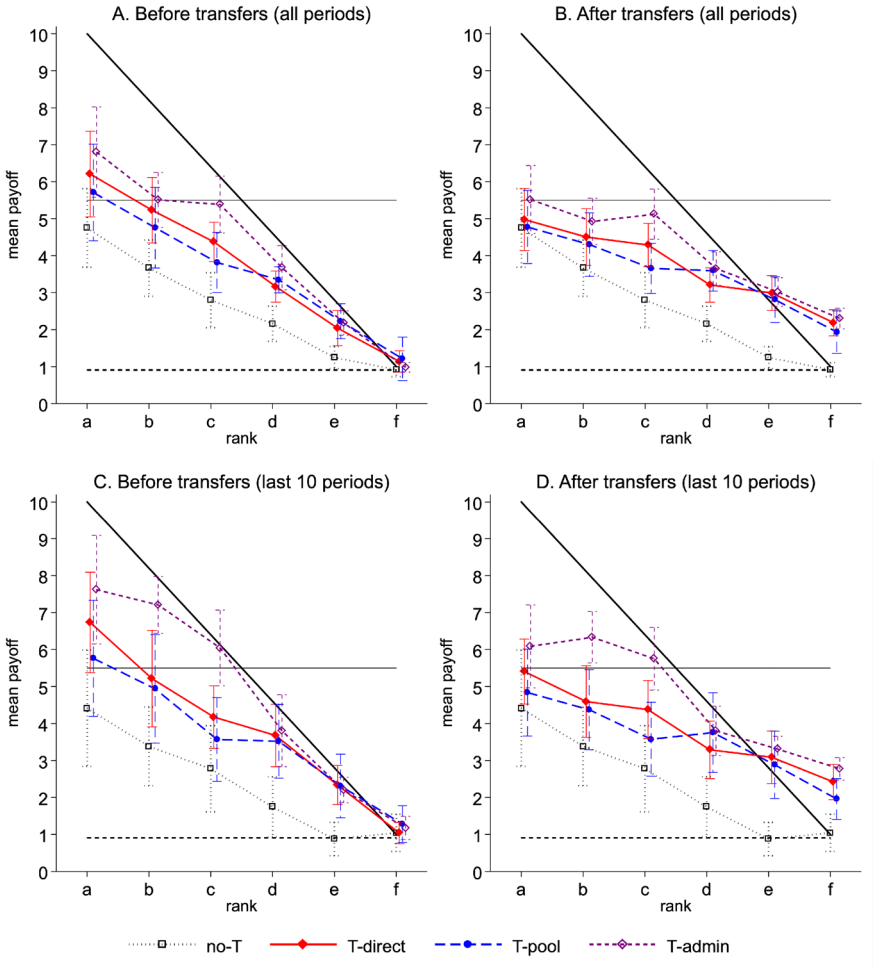


Fig. 5 Effect of redistribution on payoff, by ranks. Mean payoffs *before* (left) and *after* (right) transfers, averaged over **A**–**B** periods 1–50 or **C**–**D** periods 41–50. The dashed horizontal line at .9 denotes expected payoffs in the mixed equilibrium. The solid diagonal line shows expected payoffs in the bourgeois equilibrium without transfers. The solid horizontal line at 5.5 shows the bourgeois equilibrium with maximum volume of transfers. The 95% confidence intervals capture the between-group variation around the treatment means

being exogenously imposed rather than voluntary. Zooming in on the last 10 periods, we see that the dashed purple *T-admin* line is consistently above the solid red *T-direct* line. The main beneficiaries of that lack of discretion are in fact ranks *b* ($p = .013$) and *c* ($p = .032$).

Result 3 *The exogenous status hierarchy translates into a hierarchy of payoffs, even after transfers.*

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Result 4 *Lower ranks benefit both directly from the presence of transfer institutions—through the transfers received—and indirectly—through the reduction of miscoordination.*

4.2 Compliance with the status quo

In the bourgeois equilibrium, players are predicted to follow the correlation device and thus choose the action *claim* when the message is *claim*, and to *concede* when told to *concede*. The *critical condition* for the existence of the bourgeois equilibrium is the lowest-ranked player's belief about others' disposition to obey the recommendation of the correlation device. In the absence of transfers (Proposition 1), rank *f* needs to believe that others will obey with at least 91% probability (and disobey with 9% probability). In the presence of transfers (Proposition 2), she only needs an obedience belief of 50%. In Figure A2, we plot the relative frequency of full coordination on the bourgeois equilibrium in the last 10 periods against the mean obedience of others, as experienced by rank *f* in the first 40 periods.[30] We find that in the *no-T* treatment, 7 of 16 groups do not satisfy the condition. In the three transfer treatments (*T-direct*, *T-pool*, and *T-admin*), only 1 of 48 groups does not satisfy the critical condition. In all four treatments there is a strong positive correlation between the mean obedience experienced by rank *f* in the first 40 periods and the relative frequency of full coordination on the bourgeois equilibrium in the last 10 periods.[31]

Figure 6 shows the conditional probabilities for (a) claiming and (b) conceding in the different treatments. Across all treatments, and for all ranks, the propensity to claim when told to claim (i.e. to choose *red* when it shown in bold) is around 90%, corroborating the findings of Anbarcı et al. (2018). In contrast, the willingness to concede when told to concede (i.e. to choose *blue* when it shown in bold) appears to substantially vary with rank. In *no-T*, the higher a player's rank, the higher her willingness to concede when told to do so, ranging from 63% for rank *b* to 34% for *f* ($p = .023$).[32] This is in line with the idea that a player is more willing to abide by a given order, the higher her expected benefits from the order's existence. Or put differently, a deviation from the bourgeois equilibrium is less likely, the higher a player's expected cost from the deviation.

Yet interestingly, this logic does not seem to translate to the transfer treatments. In fact, we find the entire positive effect of transfer institutions on conceding to be driven by the bottom ranks. Compared to the *no-T* treatment, conceding of rank *f*

[30] The mean obedience of others that rank *f* experiences in her first 40 encounters is a proxy for *f*'s belief about others' disposition to obey the recommendation of the correlation device in the last 10 encounters.

[31] A simple OLS regression reveals a slope of 2.82 ($p = .033$) for *no-T*, 1.39 ($p = .014$) for *T-direct*, 1.03 ($p = .045$) for *T-pool*, and 1.92 ($p < .001$) for *T-admin*.

[32] For comparisons between different ranks of the same treatment, we report P-values of two-sided Wilcoxon signed-rank tests of matched pairs. Decisions are first averaged over 50 periods, by individual. The means of rank *i* and rank *j* are then matched by group. For each comparison of any two ranks, there are thus 16 matched-pairs per treatment.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

increases by 24 percentage points in *T-direct* and *T-pool* ($p < .033$), respectively, and by additional 20 percentage points in *T-admin* ($p = .067$). For rank *e* the increase is 32–39 percentage points ($p < .007$). For rank *d* there is a significant rise in conceding of 25 percentage points in *T-direct* and *T-admin* ($p < .008$) but no effect in *T-pool*. In contrast, the willingness to concede (i.e. to choose *blue* when it shown in bold) of ranks *b* and *c* is virtually unaffected by the presence of redistribution.[33]

Since transfers in *T-admin* and *T-direct* were virtually identical, both over ranks and over time (see Figure A5), the observed difference between both treatments reveals the importance of motives for withholding conceding beyond purely distributional concerns. As noted in Sect. 3.3, plausible arguments for *T-admin* generating higher levels of conceding could be the absence of strategic reasons to withhold conceding, and less dissatisfaction with disappointing transfers coming from an impersonal process rather than from another human being. Arguably, differences in dissatisfaction should rather play out at later stages of the game when players have learned the typical level of transfers. In contrast, if lower ranks were strategically withholding their conceding in order to squeeze out higher transfers later in the game, we should rather observe contrasting patterns in the early stages of the game.

Individual patterns of conceding over time (Figure A3), show that (a) in both treatments there is a decreasing share of individuals who—within blocks of 5 periods—play the mixed strategy *sometimes concede*. (b) In *T-direct*, there is a higher share of individuals who play the pure strategy *never concede* throughout the game ($p = .083$). (c) The share of *always conceders* increases in both treatments until period 30 but then plateaus slightly below 60% in *T-direct* while continuing to rise to 80% in *T-admin*. In the last 10 periods of the game, the share of *always conceders* in *T-admin* is significantly higher than in *T-direct* ($p = .028$). The divergence of conceding patterns in the later stages of the game suggests differences in dissatisfaction as the main behavioral channel for the higher effectiveness of *T-admin* compared to *T-direct*.

Result 5 *Exogenously administered, compulsory redistribution increases lower ranks willingness to concede stronger that discretionary transfers.*

4.3 Effect of transfers

Mean transfers are significantly larger in *T-direct* than in *T-pool* ($p = .046$) as only 10% (9 of 87) of players give exactly zero throughout the entire game in *T-direct*, compared to 26% (23 of 89) in *T-pool* ($p = .008$).[34] The gap is driven by the willingness to transfer of the lower ranks. In particular, rank *d* ($p = .002$) and *e* ($p = .044$)

[33] The treatment differences in players' willingness to concede become even more pronounced in late periods of the game. In the last 10 periods of the game, rank *f*'s willingness to concede is 21 percentage points higher ($p = .025$) when transfers (with identical transfer volume, by design) are exogenously administered (*T-admin*) instead of freely chosen by their counterpart (*T-direct*). See Figure A4.

[34] See Figures A5 and A7 for more details. All transfer figures reported in this section describe mean transfers from $x_i = 10$ to $x_i = 1$ players (and not from $x_i = 1$ to $x_i = 10$ players), which account for $3333/3417 = 97.5\%$ and $2391/2471 = 96.8\%$ of all tokens transfered in *T-direct* and *T-pool*, respectively.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

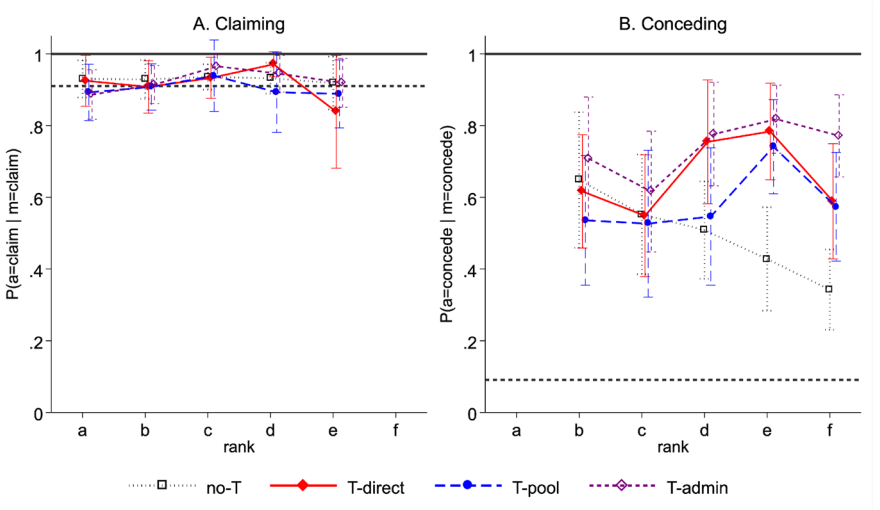


Fig. 6 Compliance with the status quo. Mean relative frequency of complying with the exogenous recommendation when one's message is **A** *claim* (i.e. red field shown in bold) or **B** *concede* (i.e. blue field shown in bold). The solid (dashed) horizontal line denotes predicted behavior in the bourgeois equilibrium (mixed equilibrium). The 95% confidence intervals capture the between-group variation around the treatment means

transfer significantly more in *T-direct* than in *T-pool*. In fact, in *T-direct* the two lower ranks give significantly higher transfers ($p = .038$) than the three upper ranks. Both the fact that transfers are higher in *T-direct* than in *T-pool*, and that—within *T-direct*—lower ranks are more generous than higher ranks, resonates with the idea that subjects conceive the decision to transfer from a backward-looking (rather than a forward-looking) perspective.[35] Transfer behavior is very stable over time.

From the perspective of the transfer recipients, transfers in *T-direct* and *T-pool* not only differ in terms of their level but also in terms of their period-by-period variation (see Figure A6). As expected, the central pooling mechanism reduces the standard deviation (SD) of received transfers from 1.27 in *T-direct* to .70 in *T-pool* ($p < .001$) as idiosyncratic differences in high-ranks' generosity are ironed out. On the other hand, received transfers in *T-admin* have a higher period-by-period variation (SD = 1.71) than in *T-direct* ($p < .001$) as transfers in all groups of *T-admin* were drawn from the *same* empirical distribution, which included the most generous and the most tightfisted individuals from *T-direct*.

In this game, transfers as such are *zero-sum* by design.[36] But indirectly they can contribute to higher overall efficiency if they lead to more conceding, and thus

[35] Note that if transfer behavior were mainly driven by "last place aversion" (Kuziemko et al. 2014), we should in fact have observed the opposite: lower ranks in *T-direct* transferring (1) less than the upper ranks in *T-direct*, and (2) less than the bottom ranks in *T-pool*.

[36] In contrast for instance to contributions in the extensively-studied public good game (Isaac et al. 1985), which are *positive-sum* by design.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

higher rates of coordination. Panel A of Fig. 7 shows that within each of the two endogenous transfer treatments *T-direct* and *T-pool*, groups with higher average transfers indeed tend to achieve a higher willingness to concede (i.e. to choose blue when blue is shown in bold).[37] A simple OLS regression on this highly aggregated data reveals an intercept of .42 ($p = .005$) and a slope of .13 ($p = .049$) for *T-direct*, as well as an intercept of .46 ($p < .001$) and a slope of .11 ($p = .071$) for *T-pool*. A group with zero transfers is thus predicted to have a willingness to concede of about 44%, which happens to correspond exactly to the mean level of conceding observed in the *no-T* treatment (see Table A1). This suggests that the mere availability of transfer opportunities has no effect on coordination but instead whether and how transfers are used. As the average transfer volume in a group increases by 1 point, conceding increases by about 12 percentage points. The same pattern holds when looking at the last 10 periods only.[38] In fact, several groups manage to converge to perfect conceding as predicted in the bourgeois equilibrium, with rather diverse levels of transfers: one group in *no-T*, two in *T-direct*, three in *T-pool*, and six in *T-admin*. The stark difference between *T-direct* and *T-admin* underlines how much more effective redistribution is when administered exogenously. And yet, in all treatments, most groups fail to reach the bourgeois equilibrium, thus foregoing the chance at substantial (material) Pareto improvements.

While the observed positive correlation between transfers and conceding is indeed consistent with our theoretical conjecture that (1) higher transfers lead to higher conceding, it could, in principle, also reflect two other causal relationships: (2) higher conceding is rewarded by higher transfers, and (3) the same individual traits determine both a player's willingness to transfer and her willingness to concede. To shed some additional light on the underlying behavioral channel, we take advantage of the panel structure of our experimental data, and of the fact that both the players' ranks as well as the matching of players into groups and encounters were determined exogenously.

Column (1a) of Table 1 presents evidence on how *lower* ranks' $(d - f)$ individual willingness to concede in the *last* 10 periods of the game can be explained by *upper* ranks' $(a - c)$ mean transfers in the *first* 10 periods of the game. For *T-direct*, we find that when upper ranks' mean transfer in the early periods was 1 token higher, the lower rank's likelihood of conceding in the late periods is about 13 percentage points higher. This holds similarly for *T-pool*. In contrast, early transfers have no impact on late conceding in *T-admin*.[39] This suggests that it is not the early transfers *per se* that explain the late conceding. Rather, early transfers in *T-direct* and *T-pool* reflect a general (i.e. home-grown) willingness to transfer of the upper ranks within

[37] Since, by design, all groups in *T-admin* drew transfers from the exact same distribution, in Fig. 7 there is virtually no variance of transfers between groups. And yet, there is substantial variance in the mean willingness to concede, ranging from about .4 to 1. This heterogeneity shows that even holding transfers constant, there are sizeable idiosyncratic differences between (groups of) individuals.

[38] Intercept of .50 ($p = .001$) and slope of .10 ($p = .081$) for *T-direct*, and intercept of .41 ($p = .001$) and slope of .19 ($p = .007$) for *T-pool*.

[39] The effect of transfers on conceding in *T-admin* (.135–.227) is insignificantly different from zero ($p = .232$). This also holds in the other specifications.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

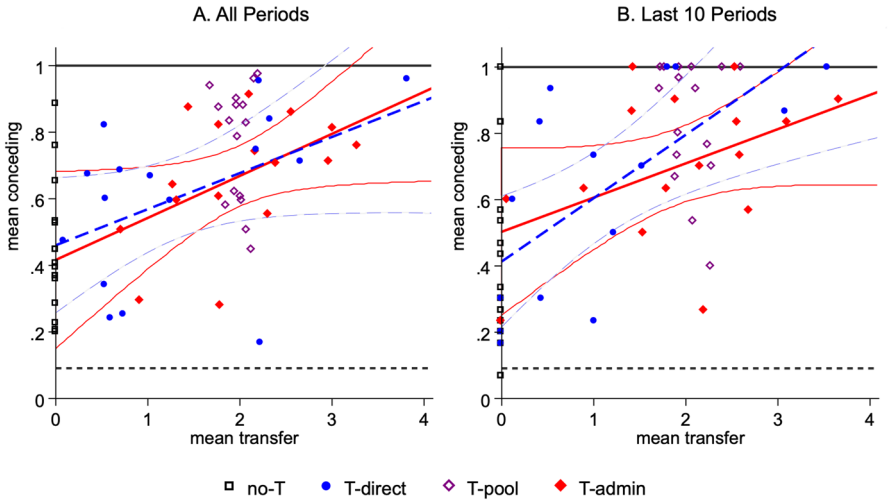


Fig. 7 Transfers and conceding, by groups. Each dot depicts one group. There are 16 groups per treatment. Mean transfer given by the player earning $x_i = h$, and $P(a = concede|m = concede)$, the mean propensity of conceding when receiving the message *concede* (i.e. choose blue when blue field is shown in bold), averaged over all ranks. **A** For all 50 periods, and **B** for the last 10 periods of the game. The solid red (dashed blue) positive-slope line depicts the simple linear regression in *T-direct* (*T-pool*), with 95% confidence intervals. Regression line for *T-admin* not included because it is insignificantly different from zero (p = .666 and p = .554). The solid (dashed) black horizontal line denotes predicted behavior in the bourgeois equilibrium (mixed equilibrium)

a given group, which lower-ranks could learn over the course of the game. In *T-admin* early transfers within a given group did, by design, not reveal anything about the level of later transfers in that particular group since the transfers of all groups were randomly drawn from the same distribution.[40] Column (1b) of Table 1 shows that the effect of transfers on conceding is robust to controlling for individual characteristics of the lower ranked players.[41] In fact, as shown in Columns (2a) and (2b), the effect is even present—albeit weaker—when restricting the explanatory variable $Transfer^{first\ \eta\ periods}_{k,\ upper}$ to the first period of the game. In period 1, *upper* ranks' transfers are uninfluenced by group dynamics, and thus exogenous to *lower* ranks' individual willingness to concede in the last 10 periods. Consequently, upper ranks' mean transfer in period 1 can be interpreted as a home-grown disposition.

Columns (3–4) of Table 1 examine the opposite causal direction, i.e. how *upper* ranks' individual willingness to concede in the *last* 10 periods of the game can be explained by *lower* ranks' mean conceding in the *first* 10 periods of the game. However, we find no evidence for a positive effect of early conceding on later transfers.[42] Instead, we find upper-ranks' transfers to be substantially driven by individuals'

[40] Figure A8 in the "Appendix" shows that groups with higher early transfers (of upper ranks) systematically display higher transfers (of upper ranks) throughout the game in *T-direct* and *T-pool* but not in *T-admin*.

[41] See "Appendix A.4", and Tables A2 and A3 for details on the individual characteristics.

[42] In fact, there is even a weakly negative effect in specification (4a) but it disappears when adding controls (4b).

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Table 1 Willingness to concede and to transfer

	DV: $Conceding_{i,\ lower}^{last\ 10\ periods}$				DV: $Transfer_{i,\ upper}^{last\ 10\ periods}$			
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
T-pool	− 0.113	− 0.125	− 0.119	− 0.159	0.056	− 0.315	0.900	0.567
	(0.168)	(0.168)	(0.169)	(0.168)	(1.194)	(1.094)	(0.951)	(0.851)
T-admin	0.574***	0.503**	0.287**	0.232*	0.870	1.066	0.011	− 0.014
	(0.198)	(0.241)	(0.120)	(0.134)	(0.961)	(0.849)	(0.396)	(0.410)
$Transfer_{k,upper}^{first\ \eta\ periods}$	0.135***	0.128**	0.101***	0.078**				
	(0.047)	(0.049)	(0.030)	(0.033)				
$Transfer_{k,upper}^{first\ \eta\ periods} \times T\text{-}pool$	0.022	0.014	− 0.008	0.001				
	(0.058)	(0.057)	(0.055)	(0.055)				
$Transfer_{k,upper}^{first\ \eta\ periods} \times T\text{-}admin$	− 0.227**	− 0.206*	− 0.097**	− 0.090*				
	(0.089)	(0.108)	(0.043)	(0.048)				
$Conceding_{k,lower}^{first\ \eta\ periods}$					0.676	0.437	− 0.867*	− 0.674
					(1.431)	(1.190)	(0.447)	(0.480)
$Conceding_{k,lower}^{first\ \eta\ periods} \times T\text{-}pool$					− 0.526	− 0.167	− 1.198	− 1.002
					(1.955)	(1.816)	(1.220)	(1.103)
$Conceding_{k,lower}^{first\ \eta\ periods} \times T\text{-}admin$					− 0.775	− 1.207	0.744	0.540
					(1.517)	(1.310)	(0.503)	(0.536)
SVO_i		0.051		0.027		0.782**		0.654**
		(0.041)		(0.046)		(0.314)		(0.273)
$Trust_i$		0.058		0.073		0.507		0.535
		(0.104)		(0.121)		(0.432)		(0.455)
$Risk_i$		− 0.042		− 0.027		− 0.139		0.013
		(0.077)		(0.072)		(0.358)		(0.347)

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Table 1 (continued)

	DV: $Conceding_{i,\,lower}^{last\ 10\ periods}$				DV: $Transfer_{i,\,upper}^{last\ 10\ periods}$			
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Age_i		− 0.006		− 0.012*		− 0.003		− 0.002
		(0.006)		(0.006)		(0.017)		(0.017)
$Female_i$		0.041		0.067		− 0.271		− 0.200
		(0.059)		(0.071)		(0.268)		(0.247)
$Siblings_i$		0.059*		0.065*		0.030		0.024
		(0.033)		(0.034)		(0.081)		(0.080)
Constant	0.476***	0.478***	0.561***	0.590***	1.055	1.282	1.925***	1.936***
	(0.127)	(0.140)	(0.097)	(0.110)	(0.893)	(0.834)	(0.362)	(0.387)
Individuals *i*	144	144	132	132	140	140	140	140
Groups *k*	48	48	44	44	48	48	48	48
R^2	0.261	0.303	0.146	0.206	0.045	0.165	0.109	0.201

OLS regression. Robust standard errors, clustered by groups, in parentheses. * ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$). Includes data from *T-pool*, *T-admin*, and *T-direct*, the latter being the reference treatment. Columns 1–2: The DV $Conceding_{i,\,lower}^{last\ 10\ periods}$ denotes the mean willingness to concede of a lower-rank player *i* of group *k* in the *last* 10 periods of the game (i.e. choose blue when blue field is shown in bold). The explanatory variable $Transfer_{k,\,upper}^{early}$ denotes the mean transfer given by the upper-rank players of group *k* in the *first* η periods, with $\eta = 10$ in Column 1 and $\eta = 1$ in Column 2. Columns 3–4: The DV $Transfer_{i,\,upper}^{last\ 10\ periods}$ denotes the mean transfer given by an upper-rank player *i* of group *k* in the *last* 10 periods of the game. The explanatory variable $Conceding_{k,\,lower}^{early}$ denotes the mean propensity of conceding by the lower-rank players of group *k* in the *first* η periods, with $\eta = 10$ in Column 3 and $\eta = 1$ in Column 4. Ranks $a - c$ ($d - f$) are defined as *upper* (*lower*) ranks. The maximum num. of obs. is 144 (3 treatments × 16 groups × 3 encounters). Observations are missing when the DV or one of the explanatory variables is not defined (e.g. when coordination failed and transfers were not possible). See "Appendix A.4", and Tables A2 and A3 for details on the individual characteristics

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Social Value Orientaion (*SVO*). *Ceteris paribus*, the transfer of a prototypically prosocial individual ($SVO = 1$) is, on average, about .7 tokens higher than of a selfish individual ($SVO = 0$). In sum, controlling for individual traits, we find evidence for higher transfers leading to higher conceding but not for higher conceding being rewarded with higher transfers.

Result 6 *Higher transfers lead to more conceding, and thus better coordination and higher efficiency.*

5 Conclusion

This paper provides first causal empirical evidence for the conjecture that redistribution can have a *positive* effect on economic efficiency by reducing disputes over individual property rights. In our experiment, redistribution not only increases overall efficiency but improves the economic status of each rank on the pre-birth status ladder. While lower ranks benefit equally from each of the three stylized transfer institutions studied, upper ranks benefit most from the setting in which the transfer decision is taken out of their hands. Typically, compulsory redistribution is justified as a means to limit temptations to free-ride on other people's charitableness. Our findings suggest an additional rationale: By mitigating the dissatisfaction from receiving disappointing transfers, non-discretionary redistribution is more effective at persuading lower ranks to obey the rules of the status quo. As a result, the main beneficiaries of exogenously administered transfers happen to be the upper ranks, who lose discretion but gain more secure claims to property. In the absence of coercive means, redistribution serves as a tool to turn privilege into economic payoff by enhancing the power of focality.

We thus add to the strand of research examining the circumstances under which redistribution enhances efficiency (e.g. Eaton and White (1991), Bowles and Gintis (1995), Birdsall et al. (1995), Midgley (1999), see Putterman et al. (1998) for an overview of the literature). We also complement a prominent strand of literature in public economics that has identified *negative* behavioral responses to redistribution, like lower effort and wasteful expenditures for avoidance. That literature typically assumes that property rights are being enforced, and focuses on individual differences in productivity and effort as drivers of inequality (Agranov and Palfrey 2015). In contrast, we assume zero coercion, and view (the absence of conflict over) property rights as the outcome of some implicit negotiation process between individuals who differ only in terms of their pre-birth privilege in the prevailing social order (status quo). To draw a more realistic picture of the net effect of redistribution on economic efficiency, further research should aim at systematically combining both stylized perspectives. Moreover, a comprehensive assessment of the merits of redistribution needs to consider its implications on expenditures for law enforcement and

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

private protection (Anderson 1999; Merlo 2003) as well as for conspicuous consumption (Hopkins and Kornienko 2004, 2009).

In our experimental setting, inequality is not inevitable. Even in the absence of transfer institutions, players could in principle reconcile efficiency and equality, for instance by following the device's recommendation in odd-numbered periods and doing the opposite in even-numbered periods.[43] The rank of a player is not *directly* payoff-relevant but it becomes *indirectly* payoff relevant through the correlation device systematically favoring higher ranks and the common expectation of compliance with the device.

In reality, of course, pre-birth status differences manifest themselves more directly through differences in resource endowments. Purely exogenous attributes like descent, primogeniture, ethnicity, and gender have historically played a major role in defining status differences concerning the access to resources, e.g. a certain piece of land, the right to exercise a certain profession, the right to vote, and other privileges or property rights (Elster 1992; Schotter and Sopher 2003; Moulin 2004). Inherited differences in endowments continue to heavily influence a person's lifetime earnings in many modern democracies, and are often reinforced by differential access to education and the ability to hedge against life risks (Bowles and Gintis 2002; Kahlenberg 2010; Chetty et al. 2014; Adermon et al. 2018). Moreover, differences in initial endowments may come along with the power to coerce others and to bend the rules of the game to one's advantage, for instance through bribes, control of media outlets and armed non-state actors (Glaeser et al. 2003; Acemoglu et al. 2015). By contrast, this paper illustrates the emergence of inequality by much softer means, namely the mere expectation of compliance with a hierarchical status quo.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2015). Democracy, redistribution, and inequality. In *Handbook of income distribution* (pp. 1885–1966). Elsevier.

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

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

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Adermon, A., Lindahl, M., & Waldenström, D. (2018). Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations. *The Economic Journal*, *128*(612), F482–F513.
Agranov, M., & Palfrey, T. (2015). Equilibrium tax rates and income redistribution: A laboratory study. *Journal of Public Economics*, *130*, 45–58.
Ahn, T., Balafoutas, L., Batsaikhan, M., Campos-Ortiz, F., Putterman, L., & Sutter, M. (2016). Securing property rights: A dilemma experiment in Austria, Mexico, Mongolia, South Korea and the United States. *Journal of Public Economics*, *143*, 115–124.
Ahn, T., Loukas, B., Batsaikhan, M., Campos-Ortiz, F., Putterman, L., & Sutter, M. (2018). Trust and communication in a property rights dilemma. *Journal of Economic Behavior and Organization*, *149*, 413–433.
Alesina, A., & Angeletos, G.-M. (2005). Fairness and redistribution. *The American Economic Review*, *95*(4), 960–980.
Alesina, A., & Giuliano, P. (2011). Preferences for redistribution. In *Handbook of social economics* (Vol. 1, pp. 93–131). Elsevier.
Alexander, R. D. (1987). *The biology of moral systems*. New Brunswick: Transaction Publisher.
Anbarcı, N., Feltovich, N., & Gürdal, M. (2018). Payoff inequity reduces the effectiveness of correlated-equilibrium recommendations. *European Economic Review*, *108*, 172–190.
Anderson, D. (1999). The aggregate burden of crime. *The Journal of Law and Economics*, *42*(2), 611–642.
Arifovic, J., & Ledyard, J. (2018). Learning to alternate. *Experimental Economics*, *21*(3), 692–721.
Aumann, R. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, *1*(1), 67–96.
Aumann, R. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, *55*, 1–18.
Baranski, A. (2016). Voluntary contributions and collective redistribution. *American Economic Journal: Microeconomics*, *8*(4), 149–173.
Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2009). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *The Economic Journal*, *120*(543), 40–79.
Bartling, B., Weber, R. A., & Yao, L. (2014). Do markets erode social responsibility? *The Quarterly Journal of Economics*, *130*(1), 219–266.
Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.
Bhaskar, V. (2000). Egalitarianism and efficiency in repeated symmetric games. *Games and Economic Behavior*, *32*(2), 247–262.
Bigoni, M., Camera, G., & Casari, M. (2019). Money is more than memory. *Journal of Monetary Economics*, *110*, 99–115.
Binmore, K. (1998). *Game theory and the social contract: Just playing* (Vol. 2). Cambridge: MIT Press.
Birdsall, N., Ross, D., & Sabot, R. (1995). Inequality and growth reconsidered: Lessons from East Asia. *The World Bank Economic Review*, *9*(3), 477–508.
Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120.
Bös, D., & Kolmar, M. (2003). Anarchy, efficiency, and redistribution. *Journal of Public Economics*, *87*(11), 2431–2457.
Bowles, S., & Gintis, H. (1995). Productivity-enhancing egalitarian policies. *International Labour Review*, *134*, 559.
Bowles, S., & Gintis, H. (2002). The inheritance of inequality. *Journal of Economic Perspectives*, *16*(3), 3–30.
Boyd, R., & Richerson, P. J. (1989). The evolution of indirect reciprocity. *Social Networks*, *11*(3), 213–236.
Camera, G., & Casari, M. (2009). Cooperation among strangers under the shadow of the future. *The American Economic Review*, *99*(3), 979–1005.
Carter, J. R., & Anderton, C. H. (2001). An experimental test of a predator-prey model of appropriation. *Journal of Economic Behavior and Organization*, *45*(1), 83–97.
Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.
Chetty, R., Hendren, N., Kline, P., & Saez, E. (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.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Cohn, A., Jessen, L., Klasnja, M., & Smeets, P. (2019). *Why do the rich oppose redistribution? An experiment with america's top 5%*. mimeo.
Crawford, V. (1982). A theory of disagreement in bargaining. *Econometrica*, *50*(3), 607.
Crawford, V., Gneezy, U., & Rottenstreich, Y. (2008). The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *The American Economic Review*, *98*, 1443–1458.
Dal Bó, E., & Dal Bó, P. (2011). Workers, warriors, and criminals: Social conflict in general equilibrium. *Journal of the European Economic Association*, *9*(4), 646–677.
Dana, J., Weber, R., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.
DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–72.
Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *The American Economic Review*, *105*(11), 3416–3442.
Diermeier, D., Egorov, G., & Sonin, K. (2017). Political economy of redistribution. *Econometrica*, *85*(3), 851–870.
Doerrenberg, P., Peichl, A., & Siegloch, S. (2017). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics*, *151*, 41–55.
Duffy, J., & Kim, M. (2005). Anarchy in the laboratory (and the role of the state). *Journal of Economic Behavior and Organization*, *56*(3), 297–329.
Duffy, J., Lai, E., & Lim, W. (2017). Coordination via correlation: An experimental study. *Economic Theory*, *64*, 265–304.
Dufwenberg, M., Gneezy, U., Güth, W., & Van Damme, E. (2001). Direct vs indirect reciprocity: An experiment. *Homo Oecon*, *18*, 19–30.
Durante, R., Putterman, L., & van der Weele, J. (2014). Preferences for redistribution and rerception of fairness: An experimental study. *Journal of the European Economic Association*, *12*(4), 1059–1086.
Durham, Y., Hirshleifer, J., & Smith, V. L. (1998). Do the rich get richer and the poor poorer? Experimental tests of a model of power. *The American Economic Review*, *88*(4), 970–983.
Eaton, B. C., & White, W. D. (1991). The distribution of wealth and the efficiency of institutions. *Economic Inquiry*, *29*(2), 336–350.
Elster, J. (1992). *Local justice: How institutions allocate scarce goods and necessary burdens*. New York: Russell Sage Foundation.
Embrey, M., Frechette, G., & Lehrer, S. (2015). Bargaining and reputation: An experiment on bargaining in the presence of behavioural types. *The Review of Economic Studies*, *82*(2), 608–631.
Engelmann, D., & Ortmann, A. (2002). *The robustness of laboratory gift exchange: A reconsideration*. Center for Economic Research and Graduate Education of Charles University, Economics Institute of the Academy of Sciences of the Czech Republic Working Paper, Prague.
Epper, T., Fehr, E., & Senn, J. (2020). *Other-regarding preferences and redistributive politics*. mimeo.
Esarey, J., Salmon, T., & Barrilleaux, C. (2012). What motivates political preferences? Self-interest, ideology, and fairness in a laboratory democracy. *Economic Inquiry*, *50*(3), 604–624.
Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, *62*(1), 287–303.
Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25.
Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181.
Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, *42*(1), 1–34.
Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.
Fetzer, T., & Marden, S. (2017). Take what you can: Property rights, contestability and conflict. *The Economic Journal*, *127*(601), 757–783.
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.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Gächter, S., & Falk, A. (2002). Reputation and reciprocity: Consequences for the labour relation. *Scandinavian Journal of Economics*, *104*(1), 1–26.
Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*(2), 169–179.
Gintis, H. (2007). The evolution of private property. *Journal of Economic Behavior and Organization*, *64*(1), 1–16.
Glaeser, E., Scheinkman, J., & Shleifer, A. (2003). The injustice of inequality. *Journal of Monetary Economics*, *50*(1), 199–222.
Greiner, B., & Levati, V. (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology*, *26*(5), 711–731.
Grossman, H. (1994). Production, appropriation, and land reform. *The American Economic Review*, *84*(3), 705–712.
Grossman, H. (1995). Robin hood and the redistribution of property income. *European Journal of Political Economy*, *11*(3), 399–410.
Grossman, H. (2001). The creation of effective property rights. *The American Economic Review*, *91*(2), 347–352.
Grossman, H., & Kim, M. (1995). Swords or plowshares? A theory of the security of claims to property. *Journal of Political Economy*, *103*(6), 1275–1288.
Grossman, H. I., & Kim, M. (1996). Predation and production. *The political economy of conflict and appropriation* (pp. 57–71).
Hamman, J., Loewenstein, G., & Weber, R. (2010). Self-interest through delegation: An additional rationale for the principal–agent relationship. *The American Economic Review*, *100*(4), 1826–46.
Hirshleifer, J. (1991). The paradox of power. *Economics and Politics*, *3*(3), 177–200.
Hirshleifer, J. (1995). Anarchy and its breakdown. *Journal of Political Economy*, *103*(1), 26–52.
Hopkins, E., & Kornienko, T. (2004). Running to keep in the same place: Consumer choice as a game of status. *The American Economic Review*, *94*(4), 1085–1107.
Hopkins, E., & Kornienko, T. (2009). Status, affluence, and inequality: Rank-based comparisons in games of status. *Games and Economic Behavior*, *67*(2), 552–568.
Isaac, M., McCue, K., & Plott, C. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, *26*(1), 51–74.
Isoni, A., Poulsen, A., Sugden, R., & Tsutsui, K. (2013). Focal points in tacit bargaining problems: Experimental evidence. *European Economic Review*, *59*, 167–188.
Kahlenberg, R. (2010). *Affirmative action for the rich: Legacy preferences in college admissions*. New York: Century Foundation Press.
Kuziemko, I., Buell, R., Reich, T., & Norton, M. (2014). Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics*, *129*(1), 105–149.
Lipsey, R., & Lancaster, K. (1956). The general theory of second best. *The Review of Economic Studies*, *24*(1), 11–32.
McCabe, K. A., Rassenti, S. J., & Smith, V. L. (1996). Game theory and reciprocity in some extensive form experimental games. *Proceedings of the National Academy of Sciences*, *93*(23), 13421–13428.
Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658–673.
Merlo, A. (2003). Income distribution, police expenditures, and crime: A political economy perspective. *Journal of the European Economic Association*, *1*(2–3), 450–458.
Midgley, J. (1999). Growth, redistribution, and welfare: Toward social investment. *Social Service Review*, *73*(1), 3–21.
Moulin, H. (2004). *Fair division and collective welfare*. Cambridge: MIT Press.
Myerson, R. (1994). Communication, correlated equilibria and incentive compatibility. *Handbook of Game Theory with Economic Applications*, *2*, 827–847.
Nowak, M., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.
Putterman, L., Roemer, J. E., & Silvestre, J. (1998). Does egalitarianism have a future? *Journal of Economic Literature*, *36*(2), 861–902.
Rapoport, A., Guyer, M., & Gordon, D. (1976). *The 2x2 game*. Ann Arbor: University of Michigan Press.
Rosenthal, R. (1989). A bounded-rationality approach to the study of noncooperative games. *International Journal of Game Theory*, *18*(3), 273–292.
Roth, A., & Murnighan, K. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, *17*(2), 189–198.

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press

Ryvkin, D., & Semykina, A. (2017). An experimental study of democracy breakdown, income and inequality. *Experimental Economics*, *20*(2), 420–447.
Saez, E., Slemrod, J., & Giertz, S. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, *50*(1), 3–50.
Sausgruber, R., & Tyran, J.-R. (2011). Are we taxing ourselves? *Journal of Public Economics*, *95*(1–2), 164–176.
Schelling, T. (1956). An essay on bargaining. *The American Economic Review*, *46*(3), 281–306.
Schelling, T. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.
Schotter, A., & Sopher, B. (2003). Social learning and coordination conventions in intergenerational games: An experimental study. *Journal of Political Economy*, *111*(3), 498–529.
Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*(1), 25–55.
Skaperdas, S. (1992). Cooperation, conflict, and power in the absence of property rights. *The American Economic Review*, *82*, 720–739.
Sonsino, D., & Sirota, J. (2003). Strategic pattern recognition—experimental evidence. *Games and Economic Behavior*, *44*(2), 390–411.
Trabandt, M., & Uhlig, H. (2011). The Laffer curve revisited. *Journal of Monetary Economics*, *58*(4), 305–327.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Konstantin Chatziathanasiou[1] · Svenja Hippel[2] · Michael Kurschilgen[3,4]

Konstantin Chatziathanasiou
kchatzia@uni-muenster.de

Svenja Hippel
svenja.hippel@uni-wuerzburg.de

[1] University of Münster, Münster, Germany

[2] University of Würzburg, Würzburg, Germany

[3] Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

[4] Max Planck Institute for Research on Collective Goods, Bonn, Germany

https://doi.org/10.1007/s10683-020-09685-5 Published online by Cambridge University Press