Spillover Effects of Institutions on Cooperative Behavior, Preferences, and Beliefs

By Florian Engl, Arno Riedl, and Roberto Weber*

Most institutions are limited in scope. We study experimentally how enforcement institutions affect behavior, preferences and beliefs beyond their direct influence over the behaviors they control. Groups play two identical public good games, with cooperation institutionally enforced in one game. Institutions generally have economically significant positive spillover effects to the unregulated game. We also observe that institutions enhance conditional cooperation preferences and beliefs about others' cooperativeness, suggesting that both factors are drivers of observed spillover effects. In additional treatments, we provide evidence for several factors, including characteristics of institutions, that enhance or limit the effectiveness and scope of spillover effects.
JEL: C92, D02, D72, H41
Keywords: Public goods, institutions, spillover effect, social preferences, beliefs

The success of any society is largely determined by the formal and informal institutions (e.g., laws and norms) that govern behavior (North, 1990; Ostrom, 1990; Acemoglu and Robinson, 2012). These institutions are often in place to overcome social dilemmas—such as public good provision, environmental protection and tax compliance—in which individual members' incentives conflict with what is best for the society at large.

Importantly, most institutions are limited in scope, and directly influence only a subset of all desirable behaviors. For example, in tax enforcement, some sources of income or wealth are quite easily monitored—and therefore subject to the relevant enforcement institutions—whereas for other forms of income authorities must rely largely on voluntary reporting. In the environmental area, some measures by governments directly restrict consumers' choices (e.g., energy saving through regulation) while changes in other domains (e.g., littering behavior) are more difficult to enforce. Therefore, an important question is whether institutions exhibit

* Engl: University of Cologne, Faculty of Management, Economics, and Social Sciences, Albertus-Magnus-Platz, 50923 Cologne, Germany (email: florian.engl@uni-koeln.de); Riedl: CESifo, IZA, Netspar & Maastricht University, Department of Microeconomics and Public Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands (email: a.riedl@maastrichtuniversity.nl); Weber: Department of Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zurich, Switzerland (email: roberto.weber@econ.uzh.ch). We are grateful for the valuable comments of the editors Johannes Hörner and Leeat Yariv and three anonymous reviewers, received on earlier versions of the paper. Florian Engl thanks the Dr. Jürgen Meyer foundation for financial support. We thank Pedro Dal Bo, Guillaume Fréchette, Bernd Irlenbusch, Martin Kocher and Bertil Tungodden for valuable discussions and participants at several seminars and conferences for helpful comments.

spillover effects, indirectly affecting behavior in areas where they do not directly apply. If that is the case, the overall effectiveness of an institution for fomenting socially beneficial behavior depends not only on how it influences behaviors that are subject to the institution, but also on how it affects behaviors beyond the institution's scope. Understanding such spillover effects to *voluntary* behavior is important for positive economics and policy.

In this paper, we use laboratory experiments to study how enforcement of pro-social behaviors by an institution in one domain spills over to another closely-related domain not governed by any institution. Existing research provides evidence that regulation and sanctions in one context can influence behavior in other settings. Our studies also provide evidence of such spillovers, but additionally document limitations to such effects and provide several novel insights into when and how spillover effects can potentially occur. We do so by using a rich design that employs a novel institutional mechanism and several measures that allow us to observe different channels—such as preferences and beliefs—through which spillover effects may be produced. Moreover, we also vary the natures of the institution—e.g., exogenous versus endogenous—and the targeted behaviors—cooperation and trust—to understand how such variation influences spillover effects.[1]

Before proceeding, we note that there are many possible ways to investigate spillovers of institutions. One could explore a context where the same people interact under the control of the institution and outside of it or a context where the regulated and unregulated behaviors involve interactions with different people. Most of our analysis focuses on the former context because, first, it fits important field situations as in our motivating examples and, second, simplifies the still relatively complex study design. Naturally, any design choice may limit the generalizability of results and whether any observed spillover effects apply to other contexts is an empirical question. We believe that the relative simplicity of our design delivers a foundation upon which it is straightforward to build more complex features. Indeed, while our focus is on spillovers involving interactions with the same group of 'partners,' our study also includes comparisons of how behaviors, preferences and beliefs directed toward 'strangers' may be differentially affected by spillover effects.

In our main study, to explore spillover effects from a regulated to an unregulated domain, we conducted experiments in which subjects simultaneously play two linear public good games in fixed four-person groups, repeatedly for 20 periods. In each period and in each of the two games, group members have fixed endowments, from which they can contribute to the group's public good. Contributions are doubled and then redistributed among all group members.

In a baseline *no institution* treatment, behavior is not governed by any enforce-

[1]We use experiments because this allows us to make causal statements about spillover effects of institutions. For correlational evidence of a relationship between institutions and pro-social behaviors outside the laboratory, see Henrich et al. (2001); Fisman and Miguel (2007); Herrmann, Thöni and Gächter (2008).

ment mechanism in either game. That is, subjects freely choose their public-good contributions and free-riding in both games is a dominant strategy for narrowly selfish agents. In the other two treatments, one of the two games is monitored by an institution (*PGG Rule*), while the other is not (*PGG No Rule*). The specific institution in our experiment corresponds to a central authority that punishes subjects who contribute less than a specified amount, the *Minimum Contribution Requirement* (*MCR*), with forfeiture of all income in that period from the respective game.[2]

The existence of spillovers and their strength may depend on the origin of the institution. For example, it has been found that democratically chosen institutions are more effective in fostering cooperation than exogenously imposed ones (Tyran and Feld, 2006; Dal Bó, Foster and Putterman, 2010; Sutter, Haigner and Kocher, 2010; Kamei, 2016). By extension, this may also hold for spillovers of institutions. We therefore vary by treatment the origin of the institution. In the *exogenous institution* treatment, the *MCR* in *PGG Rule* is exogenously imposed and set to the full endowment. In the *endogenous institution* treatment, the *MCR* is determined democratically, through unanimous consensus in a vote on a minimal *MCR* level.[3]

To provide a rich investigation of spillover effects and shed light on potential mechanisms underlying them, we elicit contribution behavior, preferences for conditional cooperation and beliefs about others' cooperation at three points in time and with respect to different reference groups, in all three treatments. First, before the 20 periods of the public good games described above, subjects play a one-shot public good game without any institution, where we elicit subjects' contribution behavior, cooperative preferences and beliefs about other subjects' cooperation, using randomly chosen subjects as the reference group. Second, immediately after the 20 periods of the repeated public good games, we again measure behavior, preferences and beliefs, now with respect to the 'partners' with whom subjects have just interacted and, thereafter, with respect to an unfamiliar group of 'strangers.' In an additional variant of our main study we replace the one-shot public good game with a one-shot trust game to investigate if the experience of an institution enforcing cooperation also affects trust, preferences for trustworthiness and beliefs about trustworthiness.

In our main study, both the *exogenous* and *endogenous* institution treatments exhibit significant positive direct effects; that is, both types of institutions strongly increase cooperation in the game where they apply, meaning that they directly influence behavior as intended. Importantly, both institutions also induce significant positive spillover effects,[4] and these are economically substantial. Together,

[2]This mechanism is similar to the minimum contribution mechanism studied by Keser, Markstädter and Schmidt (2017).

[3]Moreover, under the specific mechanism we use, which we describe in more detail later, the unique strict subgame-perfect Nash equilibrium in the voting stage is to select the same institution as the one exogenously imposed in our other treatment, that is, the *MCR* of full contributions.

[4]This is consistent with earlier research finding that prior experiences with cooperation can enhance

the direct and spillover effects lead to a substantial increase in welfare relative to the treatment without an institution. On average, the spillover effects in both treatments are similar, but they significantly differ in their dynamics. In the treatment with the endogenously adopted institution the spillover effects increase over time—consistent with the increased adoption of the institution over time—whereas they show no trend under the exogenously imposed institution. We also find that the spillover effects on contributions operate mainly through the intensive margin, leading those who already voluntarily contribute to contribute more, but have little effect in leading free-riders to cooperate. Our study is the first to separately identify these effects.

We also find that the contribution enhancing effect of an enforcement institution persists even after the institution is removed. In the unregulated one-shot public good game played after the repeated public good games described above, contributions are significantly higher when there was previously an institution in place than when there was not. This holds to similar degrees both when playing against the 'partners' with whom subjects had interacted previously and 'strangers' with whom there was no previous interaction.

Importantly, the experience with an institution also has a substantial effect on conditional cooperation preferences and beliefs about others' contributions. In comparison to the *no institution* treatment, in both treatments with an institution, beliefs about others' contributions are higher and preferences for conditional cooperation are stronger, even after the institution ceases to exist. This result again extends to strangers, with whom there was no prior interaction. To our knowledge, this is the first evidence indicating that spillover effects are due to institutions impacting both subjects' own preferences for cooperation and their beliefs about others' cooperativeness.

We also conducted additional analyses and treatments to further gain insights into the factors that influence institutional spillovers. First, we use a machine learning approach to test for possible individual heterogeneity in the degree to which institutional enforcement exhibits spillovers to unregulated behaviors. Specifically, we use causal trees to estimate heterogeneous treatment effects, based on a variety of individual characteristics that we measure at the end of the experiment. In this exploratory analysis, we find that younger individuals and those who show strong positive reciprocity exhibit smaller treatment effects. While this part of our analysis is highly exploratory and should be interpreted cautiously, it suggests that the degree to which individuals' behavior in unregulated settings may be influenced by the regulation of other actions is something that varies across individual types.

Second, to study the reasons behind the observed different dynamics of spillover effects between the exogenously imposed and endogenously adopted institutions, we conducted two further treatments. The *endogenous 0-20 institution* treatment uses a voting mechanism that facilitates the endogenous implementation of

pro-social behavior (e.g., Peysakhovich and Rand, 2016). We review this literature in the next section.

strong institutions (a *MCR* of full contribution) from the very beginning, as in the *exogenous institution* treatment. We achieve this by weakening the unanimity requirement and letting group members only vote for a *MCR* of either 0 points or full contribution (20 points). Further, we conducted an *exogenous yoked institution* treatment that exogenously implemented in each period the same *MCR* that was selected, on average, in the *endogenous institution* treatment. This treatment allows us to identify whether it is the endogeneity of the *MCR* that matters for the relatively stronger spillover effect over time observed in the *endogenous institution* treatment, or whether it is the observed temporal trend of the *MCR* levels. The results of these treatments suggest that it is the increasing trend of the *MCR*, rather than the endogenous adoption, that yields increasing spillover effects.

Third, an additional experiment tests whether the experience of an enforcement institution in a public good game produces spillover effects to trust games. Interestingly, we do do not find any such effects—high contributions in the repeated public good game do not induce more trust, greater trustworthiness or a stronger belief in the trustworthiness of others. This contrasts with earlier findings, reviewed below, demonstrating behavioral influences across games. Our results thus suggest that there are limits in the extent to which institutions exhibit spillover effects to unregulated behaviors, and that the similarity between the behaviors may play an important role.

Our findings make important contributions to a growing area of research addressing related questions. For instance, there exists experimental evidence showing that behavior or treatments in one game affect behavior in another game. This is also consistent with evidence that institutions outside the laboratory can have powerful indirect influences on behavior both inside and outside laboratory contexts (Acemoglu et al., 2003; Acemoglu and Robinson, 2006; Herrmann, Thöni and Gächter, 2008).[5] We contribute to this literature in several ways. No existing study directly investigates the spillover effects of an enforcement institution to a very similar but unregulated set of behaviors. Thus we corroborate some of the findings regarding spillovers across games using a different context that is similar to many economically relevant settings in the field involving institutional enforcement and related unregulated behaviors. This focus seems important, given the

[5]The variety of institutions to which individuals are exposed outside the laboratory also raises the important question of whether a single experience with a sanctioning institution, as in our experiment, can have a discernible spillover effect to other behaviors. This issue highlights the complexity involved in understanding the impacts of institutional spillovers and the channels through which they operate. If there are multiple such influences on behavior, then the variation in only one institution in an experiment potentially underestimates the effect that would obtain if that were the *only* institution with which individuals have experience. The fact that we—and others—observe spillovers in experiments suggests that some specific factors of an institution and the related behaviors may strengthen these indirect influences. For instance, in our main experiment the behaviors governed and ungoverned by the institution are both public goods games, which may strengthen the association as one of our additional studies suggests. Moreover, many studies document heterogeneity in early cooperativeness in social dilemma games, in the decay in cooperation in repeated games and in cooperation in one-shot games; some of this heterogeneity may reflect heterogeneous experience with institutions outside the laboratory.

ubiquity of enforcement institutions and the fact that they often govern only a subset of the socially desirable behaviors. At the same time, while we confirm that such spillovers can occur, our work also highlights that more research is needed to better understand the conditions under which they will occur.

Further, our study also investigates several possible channels through which spillovers may influence behavior. Our experiment studies both simultaneous spillovers to synchronous actions and sequential spillovers to subsequent behaviors, and we also investigate the dynamics of spillovers and whether spillovers extend to interactions with novel counterparts and to less closely related behaviors. We also separate the influence of institutional spillovers on preferences and beliefs, and investigate whether spillover effects are different for different types of individuals. These varied comparisons provide us with information both on the robustness and limits of institutional spillovers, demonstrating that the strength of spillover effects and the precise ways in which they seem to operate can vary across contexts.

The rest of the paper is organized as follows. Section I summarizes the related literature and develops hypotheses regarding the effect of institutions on behavior beyond their immediate scope. Section II introduces the experimental design of the main study, consisting of the *no institution*, *exogenous institution*, and *endogenous institution* treatments; Section III presents the results. Section IV describes the additional studies. In Section V we discuss possible explanations for our findings and Section VI concludes and suggests directions for future research.

I. Related literature and hypotheses

Prior experimental research demonstrates the effectiveness of institutions for enforcing high cooperation levels in social dilemmas (see, e.g., Ostrom, Walker and Gardner, 1992; Gächter and Fehr, 2000, for exogenously imposed institutions, and Gürerk, Irlenbusch and Rockenbach, 2006; Kosfeld, Okada and Riedl, 2009, for endogenously adopted institutions). This strand of literature studies institutions that typically work by changing the monetary incentives for non-cooperative behavior, but ignores potential spillover effects of such institutions.

Our main interest is in how the presence of an effective institution in one domain potentially extends to similar domains not governed by the institution, and whether we can identify potential channels for such spillover effects. We start from the null hypothesis of no spillover effects. This is supported under the standard preferences framework of narrow selfishness, whereby selfish players follow their dominant strategy and contribute zero in the public good game without an institution (*PGG No Rule*), independently of the presence or the type of the institution in the regulated public good game (*PGG Rule*).[6] The null hypothesis is also supported irrespective of the specific preference type (e.g., other-regarding

[6]For a detailed summary of the theoretical predictions of standard preferences, see Online Appendix A.1.

preferences) whenever subjects engage in narrow framing and make decisions in each game in isolation.

There are a number of studies that address the question of how behavior in a particular game is influenced by experience in other games.[7] Three papers are most closely related to our study. Cassar, D'Adda and Grosjean (2014) investigate how behavior in a one-shot trust game is affected by having earlier experienced one of two forms of sanctioning in a repeated prisoner's dilemma game with an outside option. They find that the sanctioning institution that induces more cooperative behavior in the prisoner's dilemma also yields more trusting and trustworthy behavior in the subsequent trust game than the less effective institution. Peysakhovich and Rand (2016) show that inducing high, rather than low, levels of cooperation in an initial finitely-repeated prisoner's dilemma game by selecting favorable payoffs and high continuation probabilities yields more pro-social behavior in subsequent public-goods, trust, dictator and ultimatum games, as well as a greater tendency to sanction non-cooperators and to agree with the belief that most people are trustworthy. Like us, they also demonstrate that such positive spillovers extend to greater cooperativeness toward new and unfamiliar counterparts and that they operate, at least partly, through beliefs about others' cooperativeness. Stagnaro, Arechar and Rand (2017) similarly find that a centralized sanctioning institution in a first-stage prisoner's dilemma game influences behavior in a subsequent dictator game, though it does not appear to affect sanctioning in a third-party punishment game.

These papers offer insights that also informed our study. For example, they provide evidence that spillovers *can* occur across games, for which our paper serves as an additional test of the robustness of such findings. We explore different kinds of institutions and behaviors and directly investigate how an enforcement institution that *partially* regulates a set of desirable actions affects behavior in the remaining set of closely related but unregulated actions. Importantly, our paper also adds to the understanding of additional channels through which spillovers might operate, about which the above studies are silent. For example, we study whether spillovers influence preferences for conditional cooperation, separately from beliefs about others' cooperativeness, and whether spillovers differentially affect the behavior of different types of individuals, including cooperators and free-

[7]For example, having subjects simultaneously play multiple games can lead them to adopt different strategies than when the games are played independently (Bednar et al., 2012; Savikhin and Sheremeta, 2013). McCarter, Samek and Sheremeta (2014) find that pro-social behavior is stronger when two simultaneously played public goods games are played with different groups than when they are played with identical groups; Falk, Fischbacher and Gächter (2013) report a small behavioral spillover between two public goods games when played in overlapping neighborhoods; Bernasconi et al. (2009) provide evidence that splitting one public good game into two simultaneously played games can increase subjects' contributions. For games played sequentially, studies show that groups that manage to sustain high efficiency levels in a weakest-link game have higher cooperation rates in subsequently played prisoner's dilemma games (Knez and Camerer, 2000); that there are learning spillovers between strategically similar games (Grimm and Mengel, 2012) and that there are modest spillovers between competitive and cooperative games that are played with the same opponents (Cason and Gangadharan, 2013). Cason, Savikhin and Sheremeta (2012) find behavioral spillovers between minimum- and median-effort coordination games when they are played sequentially, but not when they are played simultaneously.

riders. We also investigate the extent to which the dynamics of an institution and its endogeneity, when it is determined by those whose behavior it governs, matter for producing spillovers.

Our design is also inspired by several theoretical arguments that suggest how institutions can affect cooperative behavior in the unregulated *PGG No Rule*, through an effect on preferences for cooperation and through an effect on beliefs about others' cooperativeness. In the following, we summarize these arguments and the respective empirical evidence.

One line of research argues that institutional features can create a crowding-out or crowding-in of intrinsic motivations (for survey articles, see Frey and Jegen, 2001; Gneezy, Meier and Rey-Biel, 2011; Bowles and Polanía-Reyes, 2012). If institutions interact with intrinsic preferences, this interaction may spill over and influence a person's behavior even in domains in which the institution is not active (Frey, 1993). For example, exogenous control of one's behavior may compromise a sense of authority, thereby, through control aversion, leading to resistance against the behaviors an institution attempts to foment (Deci, 1975; Deci and Ryan, 1985; Deci, Koestner and Ryan, 1999; Belot and Schröder, 2016). While such resistance is not possible in domains regulated by the institution, it can manifest itself in domains beyond the institution's scope, creating negative spillover effects. It also implies that an endogenously chosen institution, where no exogenous control is exerted, might diminish or even eliminate the degree of crowding-out.

Models of pure altruism suggest that people desire certain levels of public good provision and are indifferent whether it is provided through their own or others' contributions (Bernheim, 1986; Andreoni, 1988). Therefore, such models suggest that an institution enforcing contributions in *PGG Rule* may crowd out voluntary contributions to *PGG No Rule*, one-to-one.

Conversely, institutions may increase people's intrinsic willingness to act pro-socially through the 'expressive function of laws' (Sunstein, 1996; Posner, 1997; Cooter, 1998; Benabou and Tirole, 2011). First, individuals might like to signal their law-abiding type to themselves and others. Such signaling is facilitated when institutions clearly stipulate which rules are to be followed. Second, laws can also express social norms and individuals may be motivated by a desire to follow them (Elster, 1989; López-Pérez, 2008; Krupka and Weber, 2013). This is consistent with the findings of Peysakhovich and Rand (2016), discussed above, and other evidence that policies and sanctions can affect behavior even after they are no longer in effect (d'Adda, Capraro and Tavoni, 2017; Nakashima, Halali and Halevy, 2017). For our treatments, these channels suggest that enforcement of a positive minimum contribution requirement (*MCR*) in one game (*PGG Rule*) may signal a social norm or law, with which subjects may prefer to comply even in the unregulated game (*PGG No Rule*). This implies that we should observe positive spillover effects. These effects may be stronger in the *endogenous institution* treatment, where the *MCR* is implemented by a unanimous voting process, thus potentially creating a stronger social norm.

An institution may also affect cooperative behavior in the unregulated *PGG No Rule* through beliefs about others' cooperativeness. In public good games, many people reciprocate positively to their beliefs about others' contributions (Fischbacher, Gächter and Fehr, 2001; Fehr and Fischbacher, 2004; Fischbacher and Gächter, 2010). Therefore, if institutions impact beliefs about others' cooperativeness, this would provide a channel through which cooperative behavior is affected even beyond the scope of the institution (Jehiel, 2005). Norms and laws may provide a focal point and, thus, positively affect beliefs about others' contributions (Cooter, 1998). Alternatively, observing votes for a strong institution may foster pessimistic beliefs about group members' trust and trustworthiness and, thus, also about their contributions (Sliwka, 2007).[8] Moreover, the effect on beliefs may differ between the *endogenous* and the *exogenous institution* treatment, as there is evidence that people reciprocate others' cooperation only when such cooperation is voluntary and intentional (Cettolin and Riedl, 2014). Hence, changes in beliefs may create spillover effects in either direction and they may differ across treatments.

In sum, several theoretical arguments and a few empirical studies suggest that institutions could influence behavior in unregulated environments through an effect on preferences for cooperation and beliefs about others' cooperativeness. Table 1 summarizes these arguments, the respective direction of their prediction, and whether the nature of the institution matters.

II. Main Study – Experimental Design

The experiment consists of five parts. We first describe Part II, which investigates the extent to which an institution that monitors and enforces cooperation in one environment spills over to an identical environment where the institution does not apply. The treatments in Part II vary the presence and nature of an institution. Parts I, III and IV elicit behavior, preferences and beliefs in a way that allows us to identify whether institutions have persistent influence on subjects' beliefs, preferences or both, even after the institutions cease to exist. These parts are identical across all treatments. Finally, in Part V, we measure various individual characteristics. Table 2 provides an overview of the experimental design.

A. The Game

All parts and treatments employ a linear public good game. Each of four group members ($n = 4$) is endowed with 20 points ($w = 20$) and can decide how many points to keep for him- or herself and how many to contribute to a public

[8] Using repeated prisoners' dilemma games, Galbiati, Henry and Jacquemet (2017) show that, on the one hand, past institutions can positively affect a player's current cooperativeness by increasing the belief about the partner's cooperativeness. However, on the other hand, compared to similar levels of voluntary past cooperation, institutions impede such learning about cooperativeness of others.

Table 1—: Overview of potential spillover effects through preference and belief channels

	Positive spillover effect	Negative spillover effect
Preference channels	**Expressive function of law** (see, e.g., Sunstein, 1996; Posner, 1997; Cooter, 1998) a) Laws allow to signal law-abiding type (*endo=exo*) (see, e.g., Benabou and Tirole, 2011) b) Laws signal social norms and people have a preference to follow social norms (*endo>exo*) (see, e.g., Elster, 1989; López-Pérez, 2008)	**Control aversion** Involuntary exogenous control (*exo>endo*) (see, e.g., Deci, 1975; Deci and Ryan, 1985; Deci, Koestner and Ryan, 1999) **Pure altruism** Others' contributions replace own contribution (*endo=exo*) (see, e.g., Bernheim, 1986; Andreoni, 1988)
Belief channels	**Expressive function of law** c) Laws provide a focal point (*endo=exo*) (see, e.g., Cooter, 1998)	**Signal of others' behavior** Necessity of and votes for institutions signal weak prosociality (*endo>exo*) (see, e.g., Sliwka, 2007)

Notes: (*endo>exo*) and (*exo>endo*) mean that a stronger effect is predicted in, respectively, the *endogenous institution* treatment and the *exogenous institution* treatment. (*endo=exo*) means that no difference in the effect is predicted across treatments with an institution.

good. The sum of contributions is doubled and equally distributed among all group members (implying a marginal per capita return of 0.5). Thus, given the contributions of all group members ($\mathbf{g} = (g_1, ..., g_4)$) the material payoff of group member i in the public good game equals

$$\pi_i(\mathbf{g}) = 20 - g_i + 0.5 \sum_{j=1}^{4} g_j. \tag{1}$$

Table 2—: Overview of experimental design

Part I	Behavior, preference and belief elicitation *(randomly determined group; no institution)*		
Part II	20 periods of *PGG No Rule* and *PGG Rule* *(new group - absolute stranger matching; partner matching within Part II)*		
	No institution	**Exogenous institution**	**Endogenous institution**
Part III	Behavior, preference and belief elicitation *(same group as in Part II; no institution)*		
Part IV	Behavior, preference and belief elicitation *(new group - absolute stranger matching; no institution)*		
Part V	Individual characteristics		

B. Part II: Treatment stage

At the beginning of Part II, subjects were randomly matched into four-person groups consisting of subjects who had not interacted previously. Within Part II, subjects played for 20 periods with the same group of subjects (partner matching). Part II differs between three treatments, *no institution*, *exogenous institution*, and *endogenous institution*.

No Institution treatment. — In each period of the *no institution* treatment, subjects simultaneously played two public good games with the same group members. The parameters of both games are as specified before. The two public goods games were displayed next to each other on the same computer screen. Henceforth, these games are called *PGG No Rule* and *PGG Rule*.[9] In fact, in this treatment both games were a *PGG No Rule*.

Before subjects made their contribution decisions in the two public good games, they were asked to indicate, separately and in each period, their belief about the average contribution of the other three group members in the two games. To avoid hedging, belief elicitation was not incentivized monetarily, but subjects were asked to enter their best estimates.[10] Thereafter, subjects made their contributions to the public goods. They entered and submitted them in the two games separately. They could decide to enter their contributions in either order and were free to contribute any amount in each game.

[9] In the experiment, the two games were neutrally labeled as "Task Left" and "Task Right" in all treatments.

[10] Gächter and Renner (2010) find that incentivized beliefs tend to lead to higher contribution levels in repeated public goods games than either non-incentivized beliefs or no beliefs at all. In this part of the experiment we are mainly interested in contributions and, therefore, decided against incentivizing belief elicitation.

At the end of each period, subjects were informed about the contributions of all group members to both public goods and their payoffs from both games. Contributions were displayed in descending order and it was not possible to match a group member's specific contributions across the two games. A subject's total payoff in each period consisted of the sum of the payoffs of the two games. At the end of the experiment, one of the 20 periods was randomly selected for payment. The total payoff in the randomly selected period was multiplied by 20, so that it counted for all 20 periods.[11]

Exogenous Institution treatment. — We implemented two main treatments with institutions. As in the *no institution* treatment, in both institution treatments, in each period subjects played two public good games: *PGG No Rule* and *PGG Rule*. The setup of *PGG No Rule* is identical to the games in the *no institution* treatment—i.e. subjects were free to contribute any integer amount between 0 and 20 points to the public good. The payoff structure of *PGG Rule*, however, varies by treatment. In each period *PGG Rule* is governed by an institution that monitors the group members' contributions in that game and punishes those members that contribute less than a *minimum contribution requirement* (henceforth, *MCR*). Specifically, the income from *PGG Rule* of any group member who contributes at least as many points to the group account as specified by the *MCR* is unaffected by the institution, whereas any group member who contributes fewer points to the group account than the *MCR* level forfeits any income from *PGG Rule* in that period.

In the *exogenous institution* treatment, the *MCR* in *PGG Rule* is fixed at 20, i.e., the entire endowment for that game. In each period, subjects see the *MCR* on their screen when making contribution decisions. The payoff from *PGG Rule* in the *exogenous institution* treatment is thus determined as follows:

$$\pi_i(\mathbf{g}) = \begin{cases} 20 - g_i + 0.5 \cdot \sum_{j=1}^{4} g_j & \text{if } g_i = 20, \\ 0 & \text{if } g_i < 20. \end{cases} \tag{2}$$

As in the *no institution* treatment the total per-period payoff of each subject is equal to the sum of the payoffs in *PGG No Rule* and *PGG Rule*.

Note that if one group member is penalized for contributing less than the *MCR* in *PGG Rule*, the incomes of the other group members are not affected. Thus, the other group members still benefit from any contributions made by any group member in *PGG Rule*. This reflects, for instance, an institution that confiscates part of an individual's income as a penalty.

[11] Azrieli, Chambers and Healy (2018) show that paying one randomly chosen period is incentive compatible when assuming only monotonicity.

Endogenous Institution treatment. — The *endogenous institution* treatment consists of two stages that are repeated in every period: an *institution formation stage* and a *contribution stage*. The institution again imposes a *MCR* on contributions only in *PGG Rule*. In contrast to the *exogenous institution* treatment, the *MCR* is now determined endogenously in each period in an *institution formation stage*, in which each group votes to determine the *MCR* for that group. Each group member votes for a desired *MCR* by specifying an integer value between 0 and 20. Following voting, the lowest contribution threshold that was voted for by any group member is implemented as the *MCR* for that period. We chose this particular mechanism because it imposes a unanimity requirement, in the sense that a particular *MCR* level is adopted only if everyone voted for at least that level.

At the end of the *institution formation stage*, group members were informed about the implemented *MCR* for that period and the votes that were cast. Votes were displayed in descending order and it was not possible to identify which member of the group voted for which *MCR*. After subjects were informed about the *MCR*, but before they made their contribution decisions, they were asked to indicate their belief about the other group members' average contribution to the two public good games.

The *contribution stage* is identical to the *exogenous institution* treatment, with the only difference that in each period the *MCR* is determined by group members in the *institution formation stage*. The payoff from *PGG Rule* in the *endogenous institution* treatment is thus determined as in Equation (2), but with a comparison between g_i and the endogenously implemented *MCR* replacing the comparison with 20. Payoffs and feedback were identical to those in the *exogenous institution* treatment, with the additional information of the realized *MCR* in that period.

C. Part I, III, & IV: Behavior, preference and belief elicitation

We elicited unconditional contributions, cooperative preferences and beliefs about others' cooperativeness at three points in the experiment (see Table 2). In Parts I, III and IV subjects played a one-shot four-person linear public good game with endowments of 20 and no institution. In each part, all subjects first made an unconditional contribution decision. Further, to separate beliefs and preferences, each part had two additional stages: a *belief-elicitation stage* to elicit beliefs about how much others would unconditionally contribute and a *preference-elicitation stage* to elicit contribution decisions conditional on others' contributions.

In the *belief-elicitation stage* we used the incentive-compatible Most Likely Interval elicitation rule (MLI) introduced by Schlag and van der Weele (2015). Specifically, subjects were asked to provide two integer values as the upper and the lower bound of the range of values that they believed would contain the actual average unconditional contribution of the other three group members (rounded to the nearest integer). Subjects earned 20 if they specified a range consisting of only

one number (i.e., a point prediction) and that number was equal to the actual rounded average contribution. For each unit increase in the provided range, a subject's potential earnings decreased by one point. Hence, if a subject provided an interval of width 10, then the subject earned 10 if the interval contained the actual realized average. Regardless of the width of the interval provided, subjects earned nothing if the actual average contribution of others was outside the specified range. Thus, subjects were incentivized to reveal their true beliefs as precisely as possible. The width of the range they provided is a measure of belief uncertainty.

In the *preference-elicitation stage*, we employ the strategy method for eliciting conditional contribution profiles (Selten, 1967; Fischbacher, Gächter and Fehr, 2001). After having indicated their unconditional contribution to the public good, subjects were asked to specify how much they would contribute for each of the 21 possible levels of (rounded) average contribution of the other group members. This procedure has been used in several studies to elicit conditional cooperation preferences that are independent of beliefs.

After all subjects made their unconditional and conditional contribution decisions, three of the four group members were randomly selected to implement their specified unconditional contributions. For the remaining group member, the conditional contribution decision was implemented, given the rounded average of the other group members' unconditional contributions. Subjects were paid their earnings for each of Parts I, III, and IV.

D. Part V: Individual characteristics

In Part V, we collected additional individual characteristics, which we use to explore possible heterogeneity in spillover responses. First, we elicited cognitive ability using the Cognitive Reflection Test (Frederick, 2005) and rule-following propensity via the task introduced in Erik O. Kimbrough and Alexander Vostroknutov (2017).[12] Additionally, we elicit subjects' attitudes towards risk, intertemporal discounting, altruism, reciprocity and trust. We used English translations of questions included in several waves of the German Socio Economic Panel (SOEP) survey.[13]

At the end of the experiment, subjects were asked about their age, gender, and academic major. At that point they were also asked to provide their reasoning when making the contribution decision for *PGG No Rule* and, in the *endogenous*

[12] In this task, subjects saw, on their computer screen, two baskets, one yellow and one blue, and a ball. They were told that they would earn 2 ECU if they placed the ball in the yellow basket and 1 ECU if they placed the ball in the blue basket. However, they were also informed that the 'rule' is to place the ball in the blue basket. There were no sanctions for not following the rule. This procedure was repeated for 30 balls. The number of balls placed in the blue basket provides a measure of a subject's propensity to follow an arbitrary rule at the expense of personal payoffs.

[13] All questions are reproduced in full in Online Appendix A.5. The behavioral validity of the risk and intertemporal discounting questions was established with incentivized experiments (see Dohmen et al., 2011; Vischer et al., 2013).

institution treatment, their reasoning when making the voting decision for the *MCR*.

E. General procedures

Before subjects entered the lab, they randomly drew a place card that directed them to a computer terminal where they found paper copies of the consent form and the Part I instructions. Subjects were informed that the experiment consists of several parts, but were not informed about the content of each part. At the beginning of each part, the instructions of that part were read out aloud to ensure common information regarding the content. The instructions to Part I and Part II included comprehension questions that had to be answered correctly before the respective part could begin.[14] All sessions were conducted in English.

We conducted six sessions on three consecutive days in November 2014 in Maastricht, Netherlands, with 136 subjects in total and six sessions on three consecutive days in February 2015 in Zurich, Switzerland, with 136 subjects in total. Treatments were randomized across sessions and each treatment was run four times, twice in the morning and twice in the afternoon, twice in Maastricht and twice in Zurich. Each subject participated only once. Overall, 92 subjects participated in the *no institution* treatment, 88 in the *exogenous institution* treatment, and 92 in the *endogenous institution* treatment.

The sessions in Zurich took place at the Laboratory for Behavioral and Experimental Economics at the University of Zurich and the sessions in Maastricht took place at the Behavioral and Experimental Economics Laboratory (BEElab) at Maastricht University. The experiments were run with the software "z-Tree" (Fischbacher, 2007). We used the software packages "hroot" (Bock, Baetge and Nicklisch, 2014) and "ORSEE" (Greiner, 2015) for recruitment. Subjects were students from the University of Zurich, the Swiss Federal Institute of Technology in Zurich and Maastricht University.

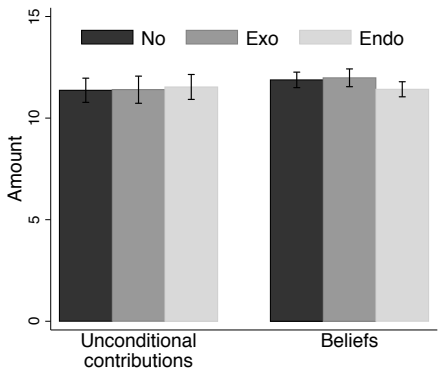
Sessions lasted about 2.5 hours. Payoffs, denominated in ECU, were converted into money at the rate of 65 ECU to €1 (about $1.25 at the time of the experiment) in Maastricht and 100 ECU to CHF 3 (about $3.25 at the time of the experiment) in Zurich. Subjects were paid anonymously at the end of the experiment. On average, subjects earned €22.52 in Maastricht, with no show-up fee, and CHF 55.45 in Zurich, which included a show-up fee of CHF 10. The total payoff from the experiment equaled the sum of the payoffs in the five parts (plus a show-up fee in Zurich). For Parts I, III, IV, and V, subjects learned about the outcomes and their payoffs only at the end of the experiment.

[14]For all treatments, the whole set of instructions together with screen shots of the decision-relevant stages can be found in Online Appendix B.

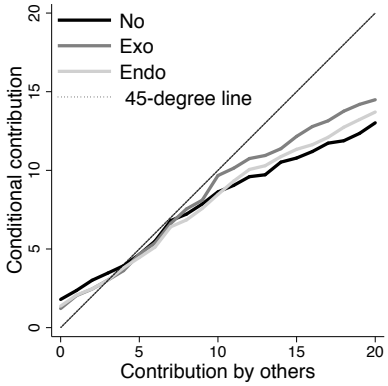
III. Results

In presenting the results, we first explore contribution behavior in Part II, during which the repeated two public good games were played and the treatments were introduced. We then jointly analyze Parts III and IV, to explore if and how the institutions introduced in Part II affect unconditional contributions, as well as cooperative preferences and beliefs about others' contributions.

Before proceeding, we make a note on behavior, beliefs and preferences measured in the one-shot game in Part I in the different treatments. Figure 1a shows the average unconditional contributions and average beliefs about others' contributions and Figure 1b the average conditional cooperation preferences, in the one-shot public good game. There are almost no differences between treatments for all three measures and statistical tests confirm that these are not significant.[15] This lack of any difference is expected, as our treatment differences were not introduced until Part II, and shows successful randomization across treatments.



a Contributions and beliefs



b Contribution preferences

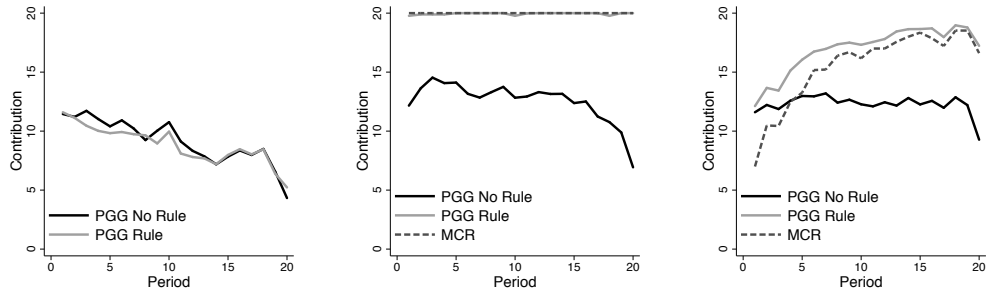
Figure 1. : Unconditional contributions, beliefs about others' contributions and cooperation preferences in Part I.

[15] These results are from pair-wise treatment comparisons of unconditional contributions (Wilcoxon rank sum tests, $p \geq 0.750$) and pair-wise treatment comparisons of beliefs (using the midpoints of elicited belief intervals) about others' contributions (Wilcoxon rank sum tests, $p \geq 0.346$). For comparing conditional cooperation preferences we construct an aggregate cooperation preference variable by taking the amount a subject decided to contribute, averaged across all possible contributions by others. Pair-wise treatment comparisons of this variable are also insignificant (Wilcoxon rank-sum tests, $p \geq 0.401$).

A. Part II contributions

Figure 2 gives an overview of behavior in the three treatments during Part II. For each treatment, the figure shows the average contributions to *PGG No Rule* and to *PGG Rule*, and—in treatments with institutions—the average *MCR*.[16]

For the *no institution* treatment, Figure 2a shows that contributions to *PGG No Rule* and *PGG Rule* closely track each other and follow the typical declining pattern found in public good games. Average contributions to *PGG No Rule* (*PGG Rule*) start at 11.45 (11.60) in the first period and decline steadily to 4.33 (5.23) in the last period. Averaged over all periods, there are no significant differences in contributions to *PGG No Rule* and *PGG Rule* (Wilcoxon signed-rank tests, $p = 0.637$).[17] Thus, when no institution was present, subjects did not behave systematically differently in the two games.



a No institution b Exogenous institution c Endogenous institution

Figure 2. : Average contributions under *PGG No Rule* and *PGG Rule*, and *MCR*.

Turning to the *exogenous institution* treatment, Figure 2b shows that this institution is—as expected—highly effective in enforcing cooperative behavior in *PGG Rule*. The *MCR* of 20 was satisfied by 1753 of 1760 contribution decisions (99.6%), yielding an average contribution to *PGG Rule* of 19.95 points. Average contributions to *PGG No Rule* start at 12.16 in the first period, slightly higher than in the *no institution* treatment, increase subsequently to reach a maximum of 14.53 in the third period, and thereafter decline steadily to 6.93 in the last period.[18]

Figure 2c shows that, in the *endogenous institution* treatment, groups do not

[16] In the *no institution* treatment, *PGG No Rule* refers to "Task Left" (i.e., the game on the left side of the decision screen) and *PGG Rule* to "Task Right" (i.e., the game on the right side).

[17] If not indicated otherwise all statistical tests are two-sided and units of observation are the independent 4-person groups. That there was little tendency to contribute differently between the two games is corroborated by group level data analysis (for details see Figure A.9(a) in Online Appendix A.2.10).

[18] At the group level, almost all groups contribute fully to *PGG Rule*, whereas there is a fair amount of dispersion in average contributions to *PGG No Rule* (see Figure A.9(b) in Online Appendix A.2.10).

immediately implement very high *MCR*'s in the *PGG Rule*. The average implemented *MCR* is 7 in the first period, but rises quickly and reaches an average of 17.75 in the final five periods. With the rise in the *MCR*, average contributions to *PGG Rule* increase over time. As in the *exogenous institution* treatment, the *MCR* in *PGG Rule* was virtually never violated (it happened only in 2 of 1840 observations). Average contributions to *PGG No Rule* start at 11.60 in the first period and remain between 11.60 and 13.20 until period 20, when they decrease to 9.27.[19]

An interesting pattern in the *endogenous institution* treatment is that groups fail to implement very high *MCR*'s, especially in the beginning, despite the fact that a *MCR* of 20 makes full contributions a strictly dominant strategy for all group members. In the first period of Part II, subjects vote, on average, for a *MCR* of 13.83, with the largest proportions voting for *MCR*'s of 20 (28.26 percent), 15 (17.39 percent) and 10 (11.96 percent).[20] The votes for relatively low *MCR*'s indicate that a sizable fraction of subjects need to learn that higher *MCR*'s are beneficial. For instance, in period 1 the actual average payoff from *PGG Rule* was 32.12 ECU and only 4 out of 92 subjects earned more than 40 ECU, the payoff with a *MCR* of 20.[21]

Figure 3 provides a closer look at the individual voting decisions over the 20 periods and shows the fraction of votes for a *MCR* of 20, a *MCR* between 10 and 19 and a *MCR* below 10. Learning is clearly visible—while a minority of subjects initially vote for a *MCR* of 20, such votes quickly increase in frequency, reaching over 80 percent after period 10. On the other hand, in every period, at least 3 subjects (out of 92) vote for thresholds below 10. Since the voting

[19] One could be concerned that the exogenously chosen high *MCR* in *PGG Rule* in the *exogenous institution* treatment creates an experimenter demand effect that leads subjects also to contribute more in *PGG No Rule*. However, in the first period, when subjects first encountered the institution, contributions to *PGG No Rule* are very similar and not statistically significant across treatments (11.45, 12.16, and 11.60, respectively, for *no*, *endogenous*, and *exogenous institutions*. Wilcoxon ranksum tests, $p > 0.488$), which speaks against such an effect.

A second concern might be that treatment effects on contributions to *PGG No Rule* are solely due to an *income effect*, due to the experimental design of Part II producing higher earnings from *PGG Rule* in the *exogenous* and *endogenous institution* treatments than under *no institution*. If this would lead to increased contributions to *PGG No Rule* through an income effect then the effect should be more pronounced in the *exogenous institution* treatment than in the *endogenous institution* treatment where the average income over all periods from *PGG Rule* was significantly lower (39.95 vs 36.95, Wilcoxon ranksum test, $p \leq 0.001$). However, this is not what we find. Over all periods, contributions to *PGG No Rule* in the *endogenous institution* treatment are not significantly different from the *exogenous institution* treatment (12.26 vs 12.53, Wilcoxon ranksum test, $p = 0.852$).

[20] Subjects' votes in period 1 correlate significantly and positively with their beliefs about group members' contributions in Part I and their own unconditional contribution decision in Part I, but not with their average conditional contributions. Furthermore, female subjects tend to vote for lower *MCR*'s (see Table A.1 in the Online Appendix).

[21] Dal Bó, Dal Bó and Eyster (2018) suggest that subjects underestimate the equilibrium effects of institutions, which can lead to demand for suboptimal policies. In our context, this could imply that subjects fail to realize that a low *MCR* will generate incentives for others to provide low contributions. However, subjects' expected payoffs from *PGG Rule*, given their contributions and their (non-incentivized) beliefs about others' contributions, from Part I are even lower (31.14 ECU) than their actual payoff. This also holds for the pivotal voters in period 1, who had no reason to update their beliefs between their vote and their contribution decision. Hence, at least these pivotal voters knowingly voted for a *MCR* that, given their stated beliefs, gave them a suboptimal payoff.

mechanism implements the minimal vote as the group's *MCR*, such subjects exert disproportionate influence. For example, in the first period, 17.39 percent of votes for a *MCR* of below 10 translate into 56.52 percent of groups with a *MCR* of below 10.

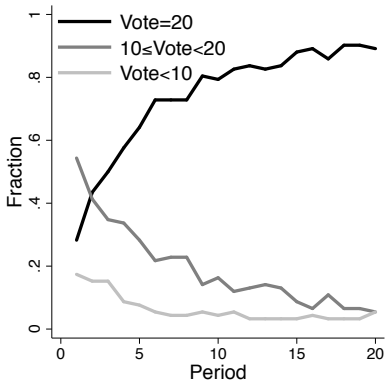


Figure 3. : Voting behavior in the *endogenous institution* treatment.

B. Direct and spillover effects of institutions

In assessing the overall effectiveness of institutions, we distinguish between direct effects and spillover effects. *Direct effects* refer to the change in contributions brought about by the institution in the game where it applies (*PGG Rule*), relative to the contributions in the respective game in the *no institution* treatment, where there is never an institution. Direct effects thus reflect the immediate influence of an institution. *Spillover effects*, in contrast, refer to the change in contributions brought about by the institution in the game where it *does not* apply (*PGG No Rule*), relative to the contributions in the respective game in the *no institution* treatment. Spillover effects thus reflect the derived or indirect effectiveness of an institution.

Direct effects. — Comparing contributions in *PGG Rule* across graphs in Figure 2 reveals that the exogenous and endogenous institutions clearly increase contributions in *PGG Rule* relative to the treatment without an institution. This direct effect is substantial and significant. Table 3 quantifies the direct effects across all periods as well as across periods 1-5, 6-10, 11-15, and 16-20. In the *exogenous institution* treatment, the direct effect amounts to 11.12 points (i.e., more than 50 percent of the endowment) over all periods and varies little over time (between 9.28 in the first five periods and 12.64 in the last five periods). In

the *endogenous institution* treatment, the direct effect increases over time from 3.48 in the first five periods to 11.02 in the last five periods, which amounts to an overall effect of 8.12. In both treatments, in all investigated blocks of periods the increase relative to the *no institution* treatment is statistically significant (Wilcoxon ranksum test, $p \leq 0.001$).[22]

Table 3—: Direct and spillover effect of institutions on contributions in the *exogenous institution* treatment and the *endogenous institution* treatment

		Period				
		All	1-5	6-10	11-15	16-20
	Direct effect	11.12	9.28	10.31	12.24	12.64
	(s.e.)	(0.02)	(0.07)	(0.05)	(0.00)	(0.05)
	[p-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Exo	Spillover effect	3.38	2.54	2.94	4.92	3.12
(N=22)	(s.e.)	(0.95)	(0.83)	(1.17)	(1.28)	(1.12)
	[p-value]	[0.026]	[0.047]	[0.073]	[0.013]	[0.052]
	Relative size of spillover effect	0.30	0.27	0.29	0.40	0.25
	Direct effect	8.12	3.48	7.53	10.47	11.02
	(s.e.)	(0.61)	(0.79)	(0.83)	(0.66)	(0.64)
	[p-value]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
Endo	Spillover effect	3.12	1.09	2.46	4.29	4.63
(N=23)	(s.e.)	(1.02)	(0.84)	(1.13)	(1.32)	(1.31)
	[p-value]	[0.047]	[0.386]	[0.153]	[0.020]	[0.013]
	Relative size of spillover effect	0.38	0.31	0.33	0.41	0.42

Notes: For each group in the treatments, the direct (spillover) effect equals the difference between that group's average contributions to *PGG Rule* (*PGG No Rule*) and the average contributions of all groups to *PGG Rule* (*PGG No Rule*) in the *no institution* treatment. p-values are from Wilcoxon rank-sum tests comparing the group-level contributions in *PGG Rule* (*PGG No Rule*) of the *exogenous* and the *endogenous institution* treatment, respectively, to contributions in *PGG Rule* (*PGG No Rule*) in the *no institution* treatment. Relative size of spillover effect is the quotient of spillover and direct effect.

[22] In all statistical tests regarding direct and spillover effects we compare the group averages in Task Left (*PGG No Rule*) and Task Right (*PGG Rule*) in the *no institution* treatment with those in the respective tasks in the *exogenous* and *endogenous institution* treatment.

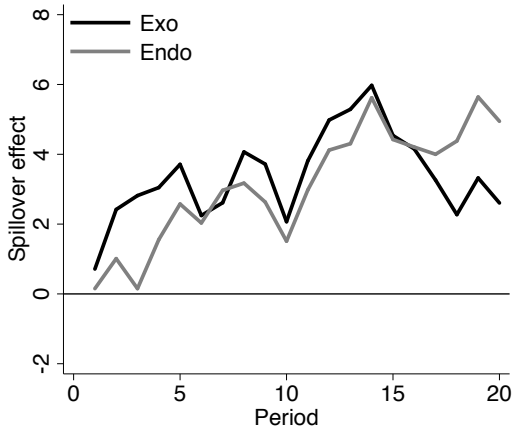


Figure 4. : Spillover effect in *PGG No Rule*.

Spillover effects. — We next investigate whether these direct effects spill over to the domain where there is no institution. Figure 4 visualizes for both treatments the spillover effects over time and Table 3 reports them for all periods and for blocks of five periods.[23] Over all periods, spillover effects are significantly larger than zero in both the *exogenous* and *endogenous institution* treatments (3.38 and 3.12, respectively; Wilcoxon rank-sum tests, $p = 0.026$ and $p = 0.047$, respectively), and there is no significant difference between these treatments (Wilcoxon rank-sum test, $p = 0.856$). (We discuss the apparently different dynamics below.)

The observed spillover effects are not only statistically significant but are also large in magnitude. Table 3 reports the relative size of the spillover effects defined as the spillover effect divided by the direct effect. Taken over all periods the relative spillover effect amounts to 0.30 and 0.38 in the *exogenous* and *endogenous institution* treatments, respectively. Thus, an institution that directly increases contributions by 100 percent, additionally increases contributions by at least 30 percent in the domain beyond its direct control. The relative spillover effects do not differ significantly between treatments, neither over all periods ($p = 0.482$) nor for any of the investigated subsets of periods ($p \geq 0.208$). Thus, our first result qualitatively reproduces earlier findings in different settings that experience with an institution can produce spillovers to related behaviors (Cassar, D'Adda and

[23] An alternative way of measuring the direct effect, taking differences in the strength of the institution (i.e., the *MCR*) into account, is reported in Table A.3 in Online Appendix A.2. There, a MCR effect is defined as the difference between that group's MCR in *PGG Rule* and the average contributions of all groups to *PGG Rule* in the *no institution* treatment. The relative size of the spillover effect is then defined as the quotient of spillover and MCR effect. When measured this way, in both treatments, the direct effects are again statistically significant and the relative spillover effects large in magnitude over all rounds and for all subsets of rounds considered.

Grosjean, 2014; Peysakhovich and Rand, 2016).

RESULT 1: *In both the* exogenous institution *treatment and the* endogenous institution *treatment, the presence of an institution that directly increases contributions in one domain induces a significantly positive spillover effect. That is, it leads to significantly increased contributions in the domain beyond the reach of the institution.*

Welfare. — Both effects contribute to increased overall welfare. Relative to the *no institution* treatment, the significant overall welfare gain in the *exogenous institution* treatment amounts to 14.50 additional points, of which 11.12 are due to the direct effect and 3.38 are due to the spillover effect (Wilcoxon rank-sum test, $p < 0.001$). Similarly, relative to the *no institution* treatment, the significant overall welfare gain in the *endogenous institution* treatment amounts to 11.24 additional points, of which 8.12 are due to the direct effect and 3.12 are due to the spillover effect (Wilcoxon rank-sum test, $p < 0.001$).

Dynamics of spillover effects. — We have seen that, taken over all periods, the average spillover effects do not differ between the *exogenous* and the *endogenous institution* treatment. However, Table 3 and Figure 4 indicate that the dynamics of the spillover effects differ across treatments. Specifically, the spillover induced by the *endogenous institution* treatment is relatively small in the beginning, but strongly increases over time, whereas the spillover induced by the *exogenous institution* treatment does not show such a clear trend. Consequently, in comparison to the *endogenous institution* treatment, the spillover effect is stronger in the *exogenous institution* treatment in earlier periods, but weaker towards the end of the 20 periods. The different dynamics also translate into different levels of statistical significance.[24,25]

RESULT 2: *The dynamics of spillover effects differ between treatments. The spillover effect from the endogenously adopted institution increases over time,*

[24] According to Spearman rank order correlations the spillover effect in the *exogenous institution* treatment does not change significantly over time ($\rho = 0.344$, $p = 0.137$), whereas the spillover effect in the *endogenous institution* treatment exhibits a significantly positive trend ($\rho = 0.901$, $p < 0.001$). The separate statistics for blocks of five periods reported in Table 3 draw a similar picture. In the *exogenous institution* treatment the spillover effect is (marginally) significantly positive in all four blocks (Wilcoxon rank-sum tests, $p \leq 0.073$) and does not change much from periods 1-5 (2.54) to periods 16-20 (3.12). In contrast, in the *endogenous institution* treatment the spillover effect increases from 1.09 in periods 1-5 to 4.63 in periods 16-20 and is not statistically significantly different from zero in the first half of the 20 periods (Wilcoxon rank-sum tests, $p \geq 0.153$), but becomes highly significant in the second half (Wilcoxon rank-sum tests, $p \leq 0.020$). The reported significance and dynamics of direct and spillover effects is corroborated by regression analysis (see Table A.2 in Online Appendix A.2.2.)

[25] Figure 4 shows a decrease of spillovers in both institution treatments in period 20. However, this decrease is small and the spillover effect in period 20 is not smaller than in a number of earlier periods in the *exogenous institution* treatment and larger than in most other periods in the *endogenous institution* treatment. In both treatments, the spillover effect in period 20 is also significantly larger than zero (*exogenous institution*: $p = 0.076$, *endogenous institution*: $p = 0.004$).

whereas the spillover effect from the exogenously imposed institutions does not show a clear trend.

It seems reasonable that the different spillover dynamics in the *exogenous* and the *endogenous institution* treatments are related to the differences in the development of the *MCR*. In the *exogenous institution* treatment, the *MCR* is fixed at 20 throughout the experiment, whereas in the *endogenous institution* treatment it increases monotonically from a rather low level in the early periods to close to 20 in the later periods. Indeed, taking group averages over all periods, contributions to *PGG No Rule* and the *MCR* in *PGG Rule* are highly correlated (Spearman's rank order correlation, $\rho = 0.315$, $p = 0.009$).[26] This suggests that there is a positive relationship between the strength of the institution and the spillover effect it creates.[27]

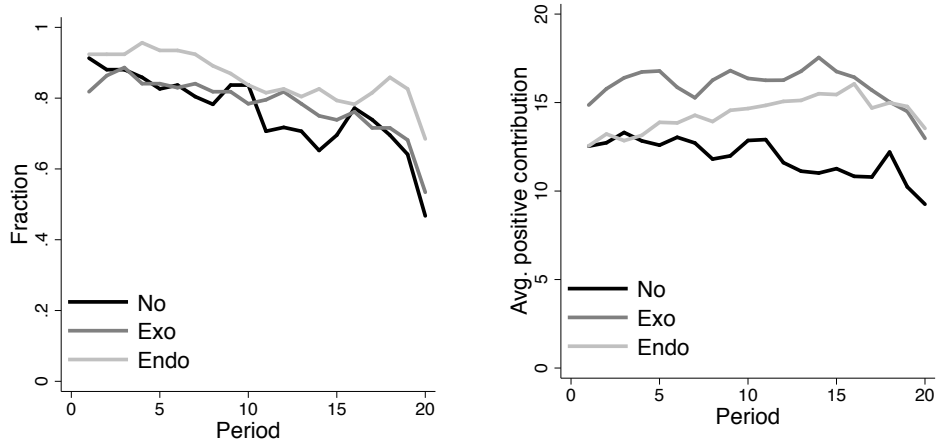
Decomposing the spillover effects. — An important remaining question is whether the observed spillover effects stem primarily from free-riders who do not contribute in the *no institution* treatment but start to contribute positive amounts when there is an institution regulating the other game, or because subjects who already contribute in the *no institution* treatment contribute more when there is an institution present, or both. We explore this by studying the spillover effects on the extensive and intensive margins of contributions.

For each treatment, Figure 5a reports the fraction of subjects who contribute a positive amount in *PGG No Rule*. It shows that there are little differences across treatments. Averaged over all periods, the frequency of positive contributions is 76.25 percent in the *no institution* treatment, 78.18 percent in the *exogenous institution* treatment, and 85.76 percent in the *endogenous institution* treatment. The differences are insignificant between the former two (Wilcoxon

[26]This is confirmed in regression analysis (see Table A.4 in Online Appendix A.2.4) that disaggregates the period-level data within a group. However, the regression results should be interpreted with caution as, in the *endogenous institution* treatment, behavior in past periods might influence both the implemented *MCR* and contributions to *PGG No Rule*.

[27]While it is not possible to completely rule out that this is driven by selection, we provide evidence against selection effects by looking at two extreme cases of *MCR*'s for which we have observations of randomly selected groups and groups that implemented the respective *MCR* endogenously. First, in the *endogenous institution* treatment we only consider groups that implemented a *MCR* of zero and compare their contributions to groups in the *no institution* treatment, in which there is also an implicit *MCR* of zero. We obtain the data for the *endogenous institution* treatment by averaging a group's contribution to *PGG No Rule* over all periods in which that group implemented a *MCR* of zero. In total, six groups implemented a *MCR* of zero at least once. We find that contributions in the former and latter case are not significantly different (average contributions to *PGG No Rule*: 7.64 and 9.15, respectively; Wilcoxon ranksum test, $p = 0.572$). Second, we look at the 19 groups in the *endogenous institution* treatment that implemented a *MCR* of 20 in some periods and compare their contributions to the *PGG No Rule* in those periods to the contributions to the *PGG No Rule* in the *exogenous institution* treatment. The average contributions in the former and the latter case are virtually identical (12.34 vs. 12.53; Wilcoxon ranksum test, $p = 0.875$). Hence, at least in these two comparisons, groups that adopted a *MCR* endogenously do not behave differently than groups for which that *MCR* was exogenously imposed. Another possibility could be that the *MCR* influences contributions in *PGG No Rule* because subjects simply imitate their contributions to *PGG Rule* in *PGG No Rule*. However, we don't find this to be the case. For a detailed analysis, see Online Appendix A.2.5.

ranksum test, $p = 0.991$), while in the *endogenous institution* treatment the fraction is marginally significantly higher (Wilcoxon ranksum tests, $p = 0.098$ and $p = 0.108$, respectively). The frequencies of positive contributions exhibit decreasing trends and Spearman's rank order correlations show that the trend is significant and similar in all treatments (*no*: $\rho = -0.903$, *exogenous*: $\rho = -0.927$, *endogenous*: $\rho = -0.828$; $p < 0.001$).



a Decision to contribute positive amount

b Level of contribution conditional on positive contribution

Figure 5. : Decision to contribute and level of contribution conditional on positive contribution.

The picture changes when looking at average contributions conditional on having contributed a positive amount (Figure 5b). Averaged over all periods, contributions are (marginally) significantly higher in the *exogenous* (15.64) and the *endogenous institution* treatments (13.91) than in the *no institution* treatment (11.49) (Wilcoxon ranksum tests, $p = 0.001$ and $p = 0.073$, respectively). The difference between the *exogenous* and the *endogenous institution* treatment is not significant (Wilcoxon ranksum test, $p = 0.128$). There is a pronounced difference in the development of average conditional contributions over time. In the *no institution* treatment contributions significantly decline (Spearman's $\rho = -0.770$, $p < 0.001$), whereas they appear to be stable in the *exogenous institution* treatment ($\rho = -0.179$, $p = 0.450$) and are significantly increasing in the *endogenous institution* treatment ($\rho = 0.698$, $p = 0.001$).[28]

[28] We corroborate these observations in regression analysis (see Table A.6 in Online Appendix A.2.6), using a hurdle model to separately estimate the treatment effects on the decision to contribute a positive amount and the conditional size of the contribution.

RESULT 3: *Decomposing the spillover effect shows that it is mainly due to an effect on the level of contributions conditional on contributing a positive amount. There is no significant effect on the decision to make a positive contribution.*

Thus, effective institutions in *PGG Rule* do not induce complete free riders to start contributing in *PGG No Rule*. Rather, the spillover effect works on those who contribute something even without enforcement institutions, by inducing them to contribute more.

C. Persistent effects of institutions on behavior, beliefs and preferences

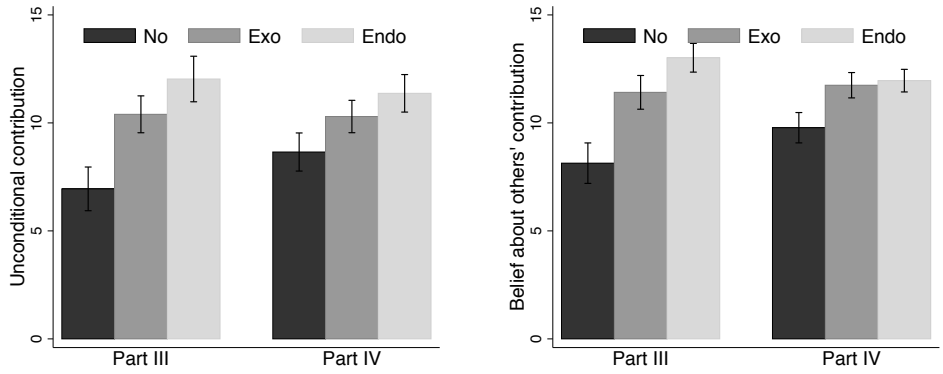
We next analyze if the experience of an enforcement institution in *PGG Rule* has a persistent effect on cooperative behavior, beliefs about others' cooperativeness, and preferences for cooperation. For this purpose, we use data from Parts III and IV, in which we elicited these variables for each subject. Recall that in these parts subjects engaged in a one-shot strategy method public goods game *without* an enforcement institution and also provided beliefs about others' cooperativeness. In Part III subjects interacted in the same group as in Part II, whereas in Part IV they interacted with subjects they had previously not encountered.

We first explore whether the effect of an enforcement institution on contribution *behavior* persists even after removal of the institution. Figure 6a shows that there are indeed clear differences in average unconditional contributions, across treatments in Parts III and IV.[29] In particular, unconditional contribution levels in the *exogenous* and *endogenous institution* treatments are significantly higher than in the *no institution* treatment Part III (Wilcoxon rank-sum tests, $p = 0.009$ and $p = 0.002$). In Part IV unconditional contributions are insignificantly higher in the *exogenous institution* treatment (Wilcoxon rank-sum tests, $p = 0.173$) and significantly higher in the *endogenous institution* treatment (Wilcoxon rank-sum tests, $p = 0.017$).[30] Our findings here are thus consistent with those observed by Cassar, D'Adda and Grosjean (2014) and Peysakhovich and Rand (2016), who report that experience with an environment promoting high levels of cooperation can increase subsequent pro-sociality.

RESULT 4: *The experience of an institution that enforces cooperation exhibits positive spillover effects not only in the simultaneous environments studied in Part II, but extends to behavior in subsequent environments without an institution. This holds also for interactions with strangers with whom there was no prior interaction.*

[29]The associated mean values, standard errors and econometric test results for contribution behavior as well as beliefs and cooperation preferences are summarized in Online Appendix A.3, Tables A.14 to A.16.

[30]In comparison to Part I contributions decrease in Parts III and IV in the *no institution* treatment (Wilcoxon signed-rank tests, $p \leq 0.002$), whereas this is not the case in the *exogenous institution* treatment and the *endogenous institution* treatment (Wilcoxon signed-rank tests, $p \geq 0.168$).



a Unconditional contribution b Belief interval - midpoint

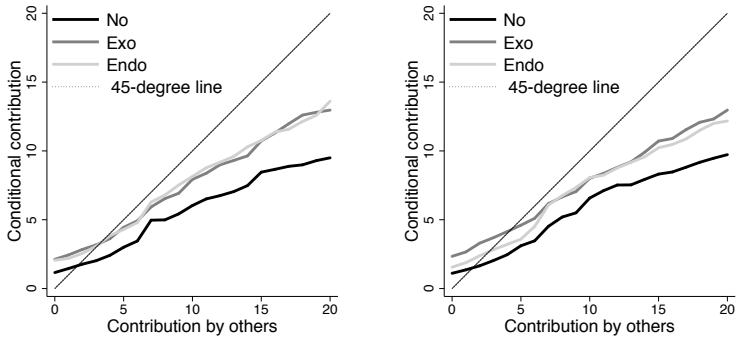
Figure 6. : Treatment effects on own contributions and beliefs about others' contributions.

The observed spillover effects could be driven by an effect on *beliefs* about others' contributions or an effect on conditional cooperation *preferences*. We first explore the role of beliefs, which were also documented in other contexts by Peysakhovich and Rand (2016). Figure 6b reveals that the average midpoints of belief intervals about others' contributions follow a pattern similar to subjects' own unconditional contributions. Statistical tests corroborate this visual impression. Beliefs about others' cooperativeness are significantly higher in the *exogenous* and *endogenous institution* treatments than under *no institution* (Part III: $p = 0.012$ and $p < 0.001$; Part IV: $p = 0.047$ and $p = 0.014$; Wilcoxon rank-sum tests). There is no significant difference between the two treatments with an institution (Wilcoxon rank-sum tests, $p \geq 0.196$). Thus, in comparison to *no institution*, experience with either the endogenously created or the exogenously imposed institution has a strong positive effect on beliefs about others' cooperativeness.[31]

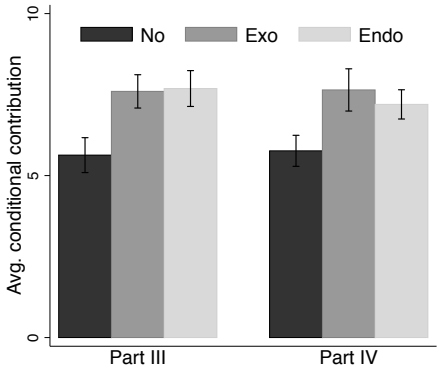
Finally, we investigate the effect of experienced institutions on cooperation *preferences*. Figures 7a and 7b show the average conditional contribution levels for each possible average contribution by the other group members in Parts III and IV, respectively. Figure 7c shows an aggregate cooperation preference variable constructed by taking the amount a subject decided to contribute, averaged across all possible contributions by others. All three figures clearly indicate stronger cooperation preferences in the treatments where subjects experienced institutions

[31] In comparison to Part I, in the *no institution* treatment, the average midpoint of the provided belief intervals is significantly lower in Part III and Part IV (Wilcoxon signed-rank test, $p < 0.001$ and $p = 0.002$), while in the *exogenous institution* treatment beliefs do not change (Wilcoxon signed-rank test, $p \geq 0.291$). In the *endogenous institution* treatment beliefs do also not change from Part I to Part IV and even increase to Part III (Wilcoxon signed-rank tests, $p = 0.260$ and $p = 0.018$).

in Part II than in the treatment without an institution.



a Part III b Part IV



c Average cooperation preferences

Figure 7. : Treatment effects on cooperation preferences.

We use the aggregate measure to test statistically for differences between treatments. The tests show that in Part III and Part IV, average cooperation preferences in the *exogenous* and the *endogenous institution* treatment are significantly stronger than in the *no institution* treatment (Wilcoxon rank-sum tests, $p \leq 0.024$). Thus, the experience of institutions in Part II not only shapes beliefs but also cooperation preferences toward others.[32]

[32]In comparison to Part I, in the *no institution* treatment average cooperation preferences strongly decrease in Part III and IV (Wilcoxon sign-rank tests, $p < 0.001$). In both treatments with institutions in Part II, cooperation preferences stay the same or decrease only weakly from Part I to Parts III and IV, respectively (*exogenous institution* treatment: $p = 0.064$ and $p = 0.127$; *endogenous institution* treatment: $p = 0.412$ and $p = 0.094$; Wilcoxon sign-rank tests). It is important to note, however, that our results cannot distinguish whether the mere experience of the sanctioning institution in PGG Rule in Part II produces spillovers to behavior in Parts III and IV, or whether such an effect operates indirectly,

RESULT 5: *In comparison to the* no institution *treatment, the experience of an institution that enforces cooperation increases beliefs about others' cooperativeness as well as preferences for cooperation, even after the institution ceases to exist. This extends to beliefs and preferences toward strangers, with whom there was no prior interaction.*

The results presented in this section have two important implications. First, institutional spillovers are not limited to concurrent decisions, but can affect behavior in subsequent games and interactions with new groups of people. Second, institutions affect beliefs about others' cooperativeness as well as own cooperation preferences, indicating that both are likely mechanisms behind the spillover effects observed in Part II.

D. Using machine learning to study heterogeneous treatment effects

In the following, we study whether the experience with our institutions affects certain individuals differently, i.e., whether there are heterogeneous treatment effects. For this we exploit individual characteristics and preferences we elicited in Part V. Knowledge about heterogeneous treatment effects can be important for policy considerations. For example, knowledge of which individuals exhibit stronger or weaker spillover effects, or even possibly negative effects, can allow policymakers to better identify the extent to which they can rely on spillover effects, rather than only on direct policy influence. We utilize machine learning tools, which are increasing in popularity in social science research (c.f. Peysakhovich and Naecker, 2017; Peysakhovich and Rand, 2017), to guide us in the identification of such effects.

More precisely, we estimate a causal tree, which creates a partition of the covariate space according to a splitting rule that optimizes for finding splits associated with treatment effect heterogeneity, and then estimates treatment effects in each element of the partition, i.e., each leaf. This allows for a visual representation of the heterogeneous effects and the associated *conditional average treatment effect* (CATE). To avoid a bias in the estimation of the sample average of the mean within a leaf, we use the honest estimation approach proposed by Athey and Imbens (2016), which splits the sample in two halves. One half of the sample is used to construct the causal tree, and the other half is used to estimate the sample means within a given leaf. Then, cross-validation techniques are used to prune the tree.[33]

through the observation of increased cooperation in PGG No Rule and a resulting impact on beliefs. We thank an anonymous reviewer for pointing this out.

[33] While causal trees are easy to interpret, it is important to keep the downsides of this method in mind. First, according to Athey and Imbens (2019), the chosen partition is somewhat arbitrary as "there may be many partitions of the data that exhibit treatment effect heterogeneity, and taking a slightly different subsample of the data might lead to a different estimated partition" (p. 711). A second concern is that when two covariates associated with strong treatment effect heterogeneity are highly correlated, only one of the two might show up as a split of the tree and the other could be disregarded, as it has no additional predictive value.

We study heterogeneous treatment effects with regard to three outcomes. First, the average contribution of a subject to *PGG No Rule* over all 20 periods, second, the unconditional contribution in Part III, and, third, the unconditional contribution in Part IV. Since only binary comparisons are possible, we consider, as treatment, only whether an individual participated in one of the treatments with an institution (i.e., in the *exogenous institution* or *endogenous institution* treatment) or in the *no institution* treatment.

Table 4—: Overview of results of the causal tree estimation

Part II - contributions to *PGG No Rule*	Part III - unconditional contributions	Part IV - unconditional contributions
Pos. reciprocity ≥ 6.5	Age < 27	Pos. reciprocity ≥ 6.5
Age < 25 | CRT ≥ 3	Pos. reciprocity ≥ 5.8 | Age ≥ 28	Age < 25 | Patience < 5

Notes: The conditional average treatment effect is smaller if the condition is fulfilled. The second line shows the split in the two subsample generated by the first split. "CRT" stands for cognitive reflection test and "Pos. reciprocity" is our survey measure of positive reciprocity, which is constructed as the average answer to three questions measured on a 7-point Likert scale. Please see the reciprocity questions (1), (4), (6) in Online Appendix A.5.

Table 4 summarizes the first two layers of the causal tree for the three outcomes.[34] Of all covariates, the two that show up as most prominent determinants of treatment effect heterogeneity are age and our survey measure for positive reciprocity. The first split in two of the trees is determined by positive reciprocity. Those with positive reciprocity above 6.5 show a smaller treatment effect than those with a lower score. In the remaining tree (for unconditional contribution in Part III), those of age below 27 show a smaller treatment effect than those above. It thus seems that especially younger cohorts and those with high intrinsic positive reciprocity are less influenced by the institution. The effect of positive reciprocity makes sense, as it is intuitive that an institution particularly influences those who have little intrinsic positive reciprocity. Moreover, it is interesting that age appears to play a role despite the relatively small variation in age in our subject pool. We also see that, on the second layer, next to age and positive reciprocity, the cognitive reflection test (CTR) score and being more or less patient also show heterogeneous treatment effects.[35]

[34]The three full trees can be found in Online Appendix A.2.7.

[35]As pointed out by an anonymous reviewer, causal trees are not the most efficient algorithms to test for heterogeneous treatment effects. Therefore, we check whether we can detect similar effects using the X-learner metaalgorithm (Künzel et al., 2019). The X-learner estimates, for each subject, a conditional average treatment effect (CATE) using all available covariates and generates confidence intervals for those individual CATEs. The fraction of significantly positive CATEs for contributions to *PGG No Rule* in Part II, unconditional contributions in Part III, and unconditional contributions in Part IV are 96.32%, 98.53%, and 50.37%, respectively. See Figure A.4 in Online Appendix A.2.7 for distributions of significant and insignificant CATEs for the three outcomes. Additionally, Figures A.5 and A.6 in Online

IV. Spillover effects: robustness tests

The treatments studied thus far show that institutions induce positive spillover effects in concurrent decisions (Part II) as well as subsequent interactions with the same and new groups of people (Part III and IV). We also provide evidence that in Part II these spillovers differ between an exogenously implemented and an endogenously adopted institution. However, this difference could not cleanly be attributed to the difference in exogeneity versus endogeneity, because in the *endogenous institution* treatment the minimum contribution requirement, *MCR*, started at a relatively low level and increased over time, whereas in the *exogenous institution* treatment the *MCR* was fixed at the maximum contribution. Therefore, it is an open question if the spillover treatment difference is due to endogeneity or due to the different time path of the strength of the institutions. Another open question relates to the fact that the domains governed and ungoverned by an institution are very similar linear public goods games. One may, therefore, ask if the observed spillover effects also exist if the domain in Parts III and IV is more distant from the linear public good in Part II. This is also of interest because earlier research documents behavioral spillovers across different domains (Peysakhovich and Rand, 2016), but further tests of the robustness of these findings are worthwhile.

To answer these questions we conducted two sets of additional treatments. We first introduce the treatments and report the results pertaining to the differences between endogenously chosen and exogenously imposed institutions. Thereafter, we discuss the treatments and results regarding whether spillover effects obtain across different domains.

A. Endogenous versus Exogenous Institutions

To disentangle whether the observed differences in our main study are due to endogeneity or to different dynamics in the strength of the institution, we conducted a follow-up study comprising two additional treatments: an *endogenous 0-20 institution* treatment and an *exogenous yoked institution* treatment. The former treatment facilitates the endogenous uptake of 20 as the implemented *MCR* right from the beginning and thus yields a cleaner comparison to the *exogenous institution* treatment. The latter implements the average *MCR* observed in the *endogenous institution* treatment exogenously for a new set of subjects and thus allows for a clean assessment of an endogeneity effect. These additional treatments

Appendix A.2.7 plot the distribution and the respective CATEs for the two individual characteristics that were found to matter most for treatment effect heterogeneity by the causal trees: age and positive reciprocity. Figures A.5(b) - A.5(d) show that, for contributions to *PGG No Rule* in Part II and the unconditional contributions in Part IV, but not those in Part III, the CATE drops visibly if positive reciprocity is above 6. This coincides with the observations from the analysis using causal trees. Figures A.6(b) - A.6(d), on the other hand, show that for unconditional contributions in Part III, those younger than age 27 have a visible smaller CATE, again as the causal tree predicted. Thus, overall, we find that the X-learner confirms what we found using causal trees.

differ from the already reported ones only in how the *MCR* is implemented in Part II. All other parts are unchanged.[36]

Design of the additional treatments. — Part II of the *endogenous 0-20 institution* treatment differs in two aspects from the original *endogenous institution* treatment. First, in each period, group members could only vote for a *MCR* of either 0 or 20. Second, the unanimity requirement was replaced by a simple majority rule; that is, a *MCR* of 20 was implemented whenever three or more out of four group members voted for it, otherwise the *MCR* was zero. As in the main *endogenous institution* treatment subjects were informed about the votes before making their contribution decisions. These changes were meant to facilitate the endogenous implementation of a *MCR* of 20. They were very effective in doing so, as the average implemented *MCR* is 18.67 and a *MCR* of 20 is implemented in 93.33% of all periods (448 out of 480).[37]

In Part II of the *exogenous yoked institution* treatment, we exogenously implemented an *MCR* in each period equal to the average *MCR* endogenously adopted in the *endogenous institution* treatment in the same period, rounded to the nearest integer. Thus, subjects in the new treatment faced, on average, the same upward trending *MCR* as subjects in the *endogenous institution* treatment (see Figure 2c).

Results. — Figure 8a shows the spillover effect in the *endogenous 0-20 institution* treatment, together with the spillover effect in the *exogenous institution* treatment.[38] The figure suggests a positive but weak spillover effect in the *endogenous 0-20 institution* treatment. Indeed, across all periods, the average spillover effect amounts to 0.71, but is not significantly different from zero (Wilcoxon ranksum test, $p = 0.595$) and is significantly smaller than the effect in the *exogenous institution* treatment (Wilcoxon ranksum test, $p = 0.036$). There is also no statistically significant time trend (Spearman's $\rho = -0.353$, $p = 0.126$).

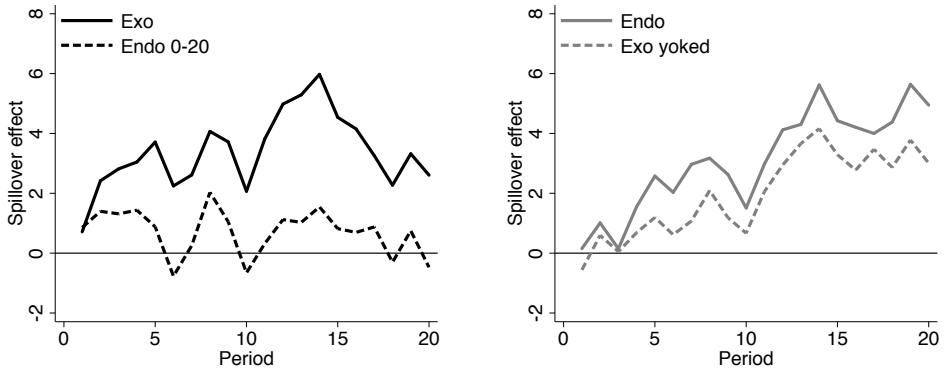
Given the significant spillover effect in the *exogenous institution* treatment, the virtual absence of it in the *endogenous 0-20 institution* treatment appears puzzling and remains an open question. In Online Appendix A.2.9 we discuss and analyze two possible intuitive explanations. First, that the small spillover effect is driven by cases where the *MCR* is zero. Second, that the deviation from

[36] We conducted additional sessions in February 2016. Only subjects that had not participated in any of the earlier sessions were invited. As before, we ran four sessions per treatment, two at Maastricht University and two at the University of Zurich. In total, we collected data from 192 additional subjects, equally distributed between the *endogenous 0-20 institution* treatment and the *exogenous yoked institution* treatment.

[37] See Figure A.7 in Online Appendix A.2.8 for the distributions of implemented *MCR*'s. There is no discernible time trend in the evolution of the *MCR* (see Figure A.8(b) in Online Appendix A.2.10).

[38] For brevity, here we report only results regarding spillover effects and relegate the results regarding direct effects to the Online Appendix; see Table A.7 and A.8 in Online Appendix A.2.9, where we also report non-parametric tests for both effects, and Online Appendix A.2.10.

unanimity in implementing the *MCR* weakens the spillover effect. We do not find significant evidence for either. However, it is clear that the endogeneity of an institution alone is not sufficient to produce strong positive spillovers.[39]



a Strong and stable institution b Institution strengthens over time

Figure 8. : Spillover effects to *PGG No Rule*.

Figure 8b shows that the spillover effect in the *exogenous yoked institution* treatment is very similar to the one in the *endogenous institution* treatment. Over all periods, the average spillover effect is somewhat weaker in the *exogenous yoked institution* treatment (1.98) than in the *endogenous institution* treatment (3.12), but the difference is statistically insignificant (Wilcoxon rank-sum test, $p = 0.444$). As in the *endogenous institution* treatment, the spillover effect in the *exogenous yoked institution* treatment exhibits a significantly increasing trend (Spearman's $\rho = 0.854$, $p < 0.001$). These results are corroborated by regression analyses (see Table A.10 in Online Appendix A.2.9).

Broadly, we interpret these results as providing suggestive evidence that the endogeneity of an institution is less relevant for the spillover effects it produces than the dynamic effects of how that institution influences behavior. Institutions that are initially weak, but strengthen over time lead to an increasing spillover effect, and this effects seems to be independent of whether the institutions are

[39]In Part I there are no significant differences between the *endogenous 0-20 institution* treatment and the *no institution* treatment, regarding unconditional contributions, beliefs about others' contributions and average cooperation preferences (Wilcoxon ranksum tests, $p \geq 0.725$). Since there is no strong spillover effect in the *endogenous 0-20 institution* treatment it is not surprising that we also do not find strong differences in unconditional contributions, beliefs and average cooperation preferences between the *endogenous 0-20 institution* treatment and the *no institution* treatment in Part III and IV, respectively (Wilcoxon ranksum tests, $p \geq 0.198$). For all non-parametric tests, see Table A.14–A.16 in Online Appendix A.3.

endogenously adopted or exogenously imposed.[40,41]

B. *Spillover effect across domains*

To explore whether the spillover effects from Part II to Parts III and IV, observed in the *endogenous institution* treatment are robust in domains different from the public good games played in Part II we conducted another two additional treatments, one equivalent to the *no institution* treatment and one equivalent to the *endogenous institution* treatment. The new treatments differ from those in the main study only in the game played in Parts I, III, and IV. In these parts, instead of playing one-shot public good games, subjects were engaged in one-shot two-person trust games (Berg, Dickhaut and McCabe, 1995). All other aspects (including the matching procedures) were the same as in the original treatments. We call these the *trust no institution* treatment and the *trust endogenous institution* treatment.[42]

Design of the additional treatments. — In Part I, each subject was endowed with 20 points and the sender in the trust game could send any portion of the endowment to the receiver who received three times the amount sent. The receiver then decided how much of the received amount to send back to the sender. As for the main study, we were interested in subjects' behavior, preferences and beliefs. Therefore, we implemented a strategy-method version of the trust game to elicit *unconditional trust*, *conditional trustworthiness*, and *beliefs about trustworthiness*. Each subject acted in both roles, sender and receiver, which was implemented in the following way.

Part I consisted of two phases. In Phase 1, each subject decided first in the role of sender how many points of the endowment to send to the receiver (eliciting

[40]Consistent with our findings for the *endogenous institution* treatment, the increasing spillover effect in the *exogenous yoked institution* treatment appears to be due to an increase of contributions among those subjects who already contributed some positive amount and not to a change in the decision whether to contribute at all (see Table A.11 in Online Appendix A.2.10).

[41]In Part I, in the *exogenous yoked institution* treatment compared to the *no institution* treatment, there are no significant differences with respect to unconditional contributions, beliefs about others' contributions and cooperation preferences (Wilcoxon ranksum tests, $p \geq 0.693$). Concerning Parts III and IV, the results in the *exogenous yoked institution* treatment are consistent with those in the *endogenous institution* treatment, but the effects are slightly weaker. In Part III, the *exogenous yoked institution* treatment leads to significantly higher unconditional contributions and beliefs about others' contributions compared to the *no institution* treatment (Wilcoxon ranksum test, $p = 0.012$ and $p = 0.029$). There is also an increase in cooperation preferences but it fails to reach statistical significance (Wilcoxon ranksum test, $p = 0.302$). In Part IV, with respect to strangers, unconditional contributions, beliefs about others' contributions and cooperation preferences are higher than in the *no institution* treatment, but the differences are not quite statistically significant (Wilcoxon ranksum tests, $p \geq 0.108$). For details, see Tables A.14–A.16.

[42]We conducted these additional sessions in November 2019. Only subjects that had not participated in any of the earlier sessions were invited. As before, we ran four sessions per treatment, two at Maastricht University and two at the University of Zurich. In total, we collected data from 184 additional subjects, with 88 participating in the *trust no institution* treatment and 96 in the *trust endogenous institution* treatment.

unconditional trust) and, thereafter, in the role of receiver how many points to send back to the sender for each possible amount the sender could send (eliciting *conditional trustworthiness preferences*). In Phase 2, each subject had to indicate their belief about how many points a receiver would send back, given a sender sent the whole endowment of 20 points (eliciting *beliefs about trustworthiness*). It was emphasized to subjects that they would not know their actual role until the end of the experiment and that they therefore should treat each role and each decision as the one that determines their earnings from this part of the experiment.[43]

Parts III and IV were exactly the same as Part I, except for the matching. In Part I, subjects were randomly matched into pairs of senders and receivers. In Part II, where they played the public good games, they were matched in groups of four as in the main study, and informed that none of the other three group members was their matched sender or receiver in Part I. In Part III, they were matched with one of the three group members with whom they interacted in Part II. Finally, in Part IV subjects were randomly assigned to a sender or receiver with whom they had not previously interacted in any of the parts. This matching protocol is analogous to the one in the main study.

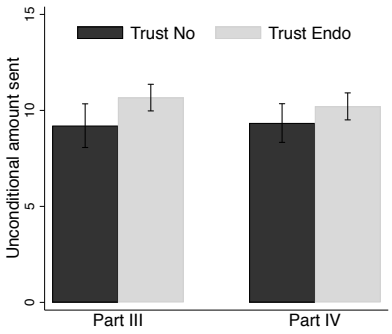
Results. — We are mainly interested in whether we observe similar spillover effects from Part II to Part III and Part IV, respectively, as in the main treatments.[44] Figures 9a–c compare unconditional trust, conditional trustworthiness and beliefs about others' trustworthiness between the *trust no institution* and the *trust endogenous institution* treatment for both Parts III and IV.

It is apparent from the figure that, in comparison to the *trust no institution* treatment, in the *trust endogenous institution* treatment there are little to no spillover effects from the experienced institution in Part II to either Part III or Part IV. There is only a small identifiable spillover effect on unconditional trust, which is statistically insignificant (Part III: $p = 0.253$; Part IV: $p = 0.481$; Wilcoxon rank-sum tests). There are also no significant differences in trustworthiness preferences (Part III: $p = 0.878$; Part IV: $p = 0.455$; Wilcoxon rank-sum tests) and beliefs about others' trustworthiness (Part III: $p = 0.613$; Part IV: $p = 0.930$; Wilcoxon rank-sum tests).[45] From this result we conclude that the

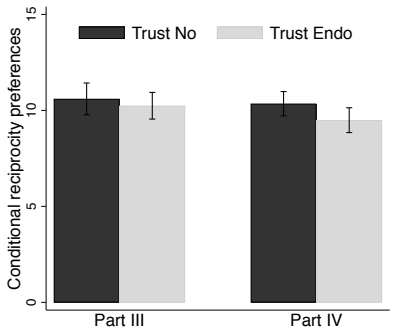
[43] All decisions were incentivized. Subjects in the role of sender received their earnings based on their unconditional trust decision and the conditional return decision of their matched receiver. Similarly, subjects in the role of receiver received their earnings based on the unconditional trust decision of their sender and their own conditional return decision. As in the other treatments belief elicitation was incentivized using the Most Likely Interval. Instructions can be found in the Online Appendix B.

[44] In Part II, we again observe a positive spillover effect, equal to about 25 percent of the direct effect, though weaker in magnitude and statistical significance, especially in later periods. Importantly, when pooling the *endogenous* and the *trust endogenous institution* treatment we find a significant spillover effect ($p = 0.015$ over all periods and $p \leq 0.038$ for periods 6–20; see Table A.18 in the Online Appendix A.4). For detailed results regarding Part II of the *trust endogenous institution* treatment, see Online Appendix A.4.

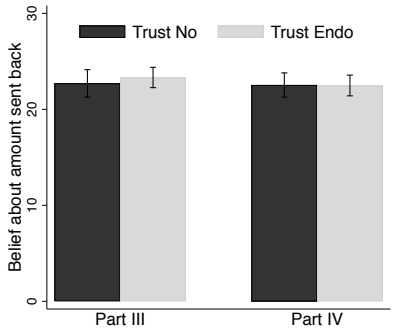
[45] In Part I, the three measures do not differ between treatments ($p \geq 0.361$; Wilcoxon rank-sum tests). Comparing Part I to respectively Parts III and IV, in each treatment, shows no difference in unconditional trust ($p \geq 0.109$; Wilcoxon signed-rank tests) and a significant decrease in both trustworthiness



a Unconditional trust



b Conditional trustworthiness



c Belief about others' trustworthiness (interval - midpoint)

Figure 9. : Treatment effects on trust, trustworthiness preferences and beliefs about others' trustworthiness.

spillover effect we observe within similar domains in the main treatments, *endogenous institution* and *exogenous institution*, do not carry over to more distant domains.

V. Discussion

Our results in the main study provide clear evidence of the existence of spillover effects from an environment governed by an institution (*PGG Rule*) to behavior in a similar environment not governed by an institution (*PGG No Rule*). We find evidence of such spillovers both when institutions are exogenously imposed and when they are endogenously determined. Moreover, our results show that these spillover effects extend to behavior, beliefs and cooperation preferences in

preferences and belief in others' trustworthiness ($p \leq 0.013$; Wilcoxon signed-rank tests); for detailed results, see Online Appendix A.4.

subsequent interactions with the same group members and with strangers. We also observe that the observed spillover effects are due both to a positive influence of institutions on beliefs regarding others' cooperativeness and on preferences for cooperation. Finally, we also document heterogeneity in the influence of spillover effects—it mainly operates by increasing the cooperativeness of those likely to exhibit some voluntary cooperation and it also appears to be related to some measurable individual characteristics, like age and a preference for reciprocity.

Incorporating our first set of additional treatments, we also find that subjects in the two treatments with an endogenously implemented institution in *PGG Rule* do not generally contribute more to *PGG No Rule* than subjects in the two treatments with an exogenously implemented institution. Thus, we cannot confirm that a spillover effect is triggered because endogenously implemented institutions signal a social norm that subjects like to follow (see the upper left column of Table 1).

Interestingly, while we observe strong and robust spillover effects on behavior, preferences and beliefs between similar decision domains, we fail to see such effects for more distant domains. Specifically, trust, trustworthiness preferences and belief in trustworthiness are not affected by the experience of institutions enforcing cooperation in public good games. This is particularly interesting, since prior studies have documented behavioral spillovers across different decision domains of social behavior (Peysakhovich and Rand, 2016). While it is beyond the scope of our paper to determine precisely why spillovers sometimes occur and do not occur in such cases, our results suggest that this is an important avenue for future research.

Viewed jointly, the above results and related earlier work suggest that the spillover effects of institutions can have substantial impacts on behavior—consistent with behavioral spillovers across games observed in earlier research—as well as preferences and beliefs, but that much remains to be done to better understand when they will occur, their strength and the factors that influence them. Our work advances this knowledge by providing additional evidence of institutional spillovers, specifically in the context of an enforcement institution, and by also documenting that belief and preference channels can both play a role. However, we also document that the existence and nature of spillover effects can be somewhat ephemeral, with conditions where we might expect a strong institutional spillover effect to obtain not producing one (as in the *endogenous 0-20 institution* treatment) and contexts in which prior studies have found spillovers (e.g., across games measuring social preferences) not producing an effect. Rather than providing a clear, simple takeaway message, our work highlights the complexity of institutional spillovers and the need to understand them better.

We can also use our evidence from the cases in which we do observe institutional spillovers to try to shed light on the possible underlying mechanisms. For instance, one potential preference mechanism is that the investigated institutions, like laws, have an expressive function and allow subjects to signal their

rule-abiding type by complying with them even when they cannot be enforced (see the upper left column of Table 1). In Part V of our experiment, using a task measuring the willingness to follow an arbitrary and unenforced rule, we elicited subjects' rule-following propensities (RFP) (cf. Kimbrough and Vostroknutov (2016)). We tested if those scoring higher on the RFP are also more willing to follow the rule highlighted by the institution in *PGG Rule*, but do not find significant effects of the interactions between subjects' rule-following propensity and any of the treatments with an institution (see column (1) of Table A.12 in Online Appendix A.2.11). This suggests that rule following is not an important reason for the observed spillover effects.

Another related potential preference channel is provided by the Social Heuristics Hypotheses (SHH), which proposes that cooperative norms establish simple heuristics that are followed intuitively. We can explore whether this channel is active using the results from the Cognitive Reflection Test (CRT), which we elicited in Part V. CRT scores are inversely related to a tendency to rely on intuition in answering questions with an immediate intuitive, but incorrect, answer. According to the SHH, if the institution in *PGG Rule* provides a heuristic for cooperation, then subjects who score lower on cognitive reflection should follow the norm intuitively and contribute more in other choices (cf. the argument in Peysakhovich and Rand (2016)). We find a negative significant effect of the interaction between the CRT score and the treatments with institutions for the *exogenous institution* treatment (see column (2) of Table A.12 in Online Appendix A.2.11). In addition, when pooling both institution treatments, our causal tree analysis in Section III.D suggests—at the second split— a stronger treatment effect in Part II for subjects with a CRT score below 3. Thus, there is some evidence—albeit weak and only suggestive—consistent with the idea that heuristic thinkers exhibit greater positive spillover effects.

We do find that institutions that strengthen over time are more successful in influencing behavior in unregulated environments than those that are strong and stable from the beginning. This is consistent with the theoretical predictions of Jackson and Acemoglu (2017) who argue that strengthening laws over time can influence social norms, independent of the mechanism with which they are implemented. These norms, in turn, might be followed also in the unregulated environment. Although the mechanism that drives this theoretical prediction—cooperation with law enforcement through whistleblowing—is not present in our experiment, it is conceivable that for other reasons social norms react more strongly to incremental increases in institutions than to large discrete changes. To investigate if this is indeed the case could be an interesting avenue for future research.

Our results in Parts III and IV of our public good treatments indicate that beliefs are a likely channel for spillover effects. Beliefs in these parts may not only reflect a pure institution effect, because they are also influenced by the history of experienced contributions in Part II. The beliefs elicited in period 1 of Part II can

provide information on the role of beliefs without this additional effect, because they were measured before any contribution decision was made or observed, but after the institution was implemented. An analysis of these beliefs shows that in the *exogenous* and in the *endogenous 0-20 institution* treatment, in period 1, beliefs about others' cooperativeness are indeed (marginally) significantly higher than in the *no institution* treatment (12.43 and 12.56 vs 11.34; Wilcoxon rank-sum tests, $p = 0.054$ and $p = 0.044$). There are no significant differences for the *endogenous* and the *exogenous yoked institution* treatment, which is likely due to the relatively low *MCR* in period 1 in these treatments. However, in the *endogenous institution* treatment there is a significantly positive correlation between the implemented *MCR* and beliefs (Spearman's $\rho = 0.279$, $p = 0.007$). Hence, the evidence from beliefs in period 1 further supports the idea that institutions in *PGG Rule* affect beliefs in *PGG No Rule*. This in turn increases contributions in *PGG No Rule*, creating at least part of the spillover effects.

A final potential mechanism operates independently of preferences and beliefs. Specifically, despite extensive experimental instructions and comprehension questions, subjects might need to learn that cooperation can provide benefits to the group and themselves. This could explain why the *exogenous institution* treatment performs better than the *endogenous institution* treatment in the beginning of Part II, as the former provides an immediate opportunity to experience the benefits of cooperation. However, we do not find support for this explanation in our data. First, if such learning is an important force, it should also be at work in the *endogenous 0-20 institution* treatment, which implemented the same strong institutions in early periods as the *exogenous institution* treatment. However, we find no significant spillover effect in the former treatment. Second, it stands to reason that such learning should i) be steepest when first witnessing the high payoffs of the strong institution, and ii) have a stronger effect for those who did worse in the unregulated compared to the regulated game. To test this we run regressions using data from the *exogenous institution* and the *endogenous 0-20 institution* treatment. We estimate the effect of the difference in profits between *PGG Rule* and *PGG No Rule* in the first period on contributions to *PGG No Rule* in the second period, while controlling for first period contributions. The regression shows a significant, but negative effect, the opposite of what we would expect if the institution would teach the subjects of the benefits of cooperation (Table A.13 in Online Appendix A.2.11). Thus, taken together, our data do not support the idea that spillover effects are due to learning that cooperation pays off.

From a broader perspective, our study can also speak to the literature on the interrelation between institutions and culture (Guiso, Sapienza and Zingales, 2006, 2008; Tabellini, 2008, 2010; Alesina and Giuliano, 2015; Gächter and Schulz, 2016). The set of beliefs and preferences that the members of a society hold are commonly acknowledged as important determinants of a society's culture. In this regard, we provide causal evidence that institutions can shape culture,

and that the institutionally induced change in culture can lead to sizable welfare effects.

VI. Conclusion

We test whether an institution that has a powerful direct effect on behavior in one domain also influences how cooperatively a group acts in a similar domain where the institution exerts no direct influence. In some of our treatment comparisons, we find clear evidence of strong spillover effects. We find that institutions that regulate cooperation in a public good game also promote cooperation in other public good games beyond the institution's scope. Treated subjects exhibit more cooperative behavior than untreated ones even after the institution is removed. Moreover, the experience with an enforcement institution leads to more positive beliefs about others' contributions and to stronger conditional cooperation preferences. These effects also extend to new counterparts, with whom no previous interaction occurred. This, together with evidence that beliefs are positively affected by the mere presence of an institution, strongly suggests that spillover effects are due to combined changes in preferences for cooperation and changes in beliefs about others' contributions. Spillover effects appear to mainly influence those people inclined to already act cooperatively, and they also heterogeneously affect different groups of people. However, we also find that the presence and size of such effects varies with the nature of the institution and the relatedness of the regulated and unregulated behaviors. We do not find spillover effects when considering spillovers from institutions regulating cooperation to more distant behaviors involving trust and trustworthiness.

While confirming that the impact of institutions *can* spill over to influence unregulated behaviors, our work also highlights that more research is necessary to understand when and how the effects of institutions extend to unregulated behaviors and settings. An obvious open question relates to the limits of institutional spillovers across domains. We investigated if enforcement in public good games spill over to trust and trustworthiness and did not find an effect. However, there are still many open questions in that respect. For instance, it remains to be seen if our null effect is robust, particularly since other research finds that inducing cooperation in one setting yields more pro-sociality in other contexts.

Moreover, even where we find positive spillovers, the generalizability of these results requires further investigation. In our study the spillover effect is measured when the institution still exists and when it just recently ceased to directly regulate behavior. However, it is an open question if a spillover effect would still be observed after some time (e.g., a week) has passed since the abolition of the enforcement institution. Further, many other institutions and settings are conceivable, and we only explore a small part of this set. For example, while our exogenously implemented institution enforced the socially most efficient cooperation level at no cost, many institutions in the field are not fully efficient or come at some other costs. This imperfection of the direct effects may limit or even

negate the positive spillovers from an institution. Future research should, therefore, examine spillover effects more broadly, across a wide variety of contexts and institutional features. Our study is only an early step in investigating the broader question of how institutions that incentivize pro-social behaviors in some domains spill over to influence pro-social behaviors in unregulated other domains.

REFERENCES

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy.* Cambridge University Press.

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* Crown Business.

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2003. "Institutional causes, macroeconomic symptoms: volatility, crises and growth." *Journal of monetary economics*, 50(1): 49–123.

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53(4): 898–944.

Andreoni, James. 1988. "Privately provided public goods in a large economy: The limits of altruism." *Journal of Public Economics*, 35(1): 57–73.

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, Susan, and Guido W Imbens. 2019. "Machine learning methods that economists should know about." *Annual Review of Economics*, 11: 685–725.

Azrieli, Yaron, Christopher J. Chambers, and Paul J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy*, 126(4): 1472–1503.

Bednar, Jenna, Yan Chen, Tracy Xiao Liu, and Scott Page. 2012. "Behavioral spillovers and cognitive load in multiple games: An experimental study." *Games and Economic Behavior*, 74(1): 12–31.

Belot, Michèle, and Marina Schröder. 2016. "The Spillover Effects of Monitoring: A Field Experiment." *Management Science*, 62(1): 37–45.

Benabou, Roland, and Jean Tirole. 2011. "Laws and Norms." *NBER Working Paper Series*, 17579.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and Economic Behavior*, 10: 122–145.

Bernasconi, Michele, Luca Corazzini, Sebastian Kube, and Michel André Maréchal. 2009. "Two are better than one! Individuals' contributions to "unpacked" public goods." *Economics Letters*, 104(1): 31–33.

Bernheim, B. Douglas. 1986. "On the voluntary and involuntary provision of public goods." *American Economic Review*, 151(3712): 789–793.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg Registration and Organization Online Tool." *European Economic Review*, 71: 117 – 120.

Bowles, Samuel, and Sandra Polanía-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50(2): 368–425.

Cason, Timothy N., and Lata Gangadharan. 2013. "Cooperation Spillovers and Price Competition in Experimental Markets." *Economic Inquiry*, 51(3): 1715–1730.

Cason, Timothy N., Anya C. Savikhin, and Roman M. Sheremeta. 2012. "Behavioral spillovers in coordination games." *European Economic Review*, 56(2): 233–245.

Cassar, Alessandra, Giovanna D'Adda, and Pauline Grosjean. 2014. "Institutional Quality, Culture, and Norms of Cooperation: Evidence from a Behavioral Field Experiment." *Journal of Law and Economics*, 57(3): 821–863.

Cettolin, Elena, and Arno Riedl. 2014. "Partial Coercion, Conditional Cooperation, and Self-Commitment in Voluntary Contributions to Public Goods." In *Coercion and Social Welfare in Public Finance: Economic and Political Dimensions.* , ed. Stan Winer and Jorge Martinez. Cambridge, UK:Cambridge University Press.

Cooter, Robert. 1998. "Expressive Law And Economics." *The Journal of Legal Studies*, 27(S2): 585–607.

d'Adda, Giovanna, Valerio Capraro, and Massimo Tavoni. 2017. "Push, don't nudge: behavioral spillovers and policy instruments." *Economic Letters*, 154: 92–95.

Dal Bó, Ernesto, Pedro Dal Bó, and Erik Eyster. 2018. "The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects." *Review of Economic Studies*, 85(2): 964–998.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100(5): 2205–2229.

Deci, Edward L. 1975. *Intrinsic Motivation.* New York:Plenum Press.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior.* New York:Plenum Press.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analyitc Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125(6): 627–628.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association*, 9(3): 522–550.

Elster, Jon. 1989. "Social Norms and Economic Theory." *The Journal of Economic Perspective*, 3(4): 99–117.

Falk, Armin, Urs Fischbacher, and Simon Gächter. 2013. "Living in Two Neighborhoods-Social Interaction Effects in the Laboratory." *Economic Inquiry*, 51(1): 563–578.

Fehr, Ernst, and Urs Fischbacher. 2004. "Social norms and human cooperation." *Trends in Cognitive Sciences*, 8(4): 185–90.

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

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review*, 100(1): 541–556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative ? Evidence from a public goods experiment." *Economics Letters*, 71(3): 397–404.

Fisman, Raymond, and Edward Miguel. 2007. "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets." *Journal of Political Economy*, 115(6): 1020–1048.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *The Journal of Economic Perspectives*, 19(4): 25–42.

Frey, Bruno S. 1993. "Motivation as a limit to pricing." *Journal of Economic Psychology*, 14(4): 635 – 664.

Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys*, 15(5): 589–611.

Gächter, Simon, and Elke Renner. 2010. "The effects of (incentivized) belief elicitation in public goods experiments." *Experimental Economics*, 13(3): 364–377.

Gächter, Simon, and Ernst Fehr. 2000. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90(4): 980–994.

Gächter, Simon, and Jonathan Schulz. 2016. "Intrinsic Honesty and the Prevalence of Rule Violations across Societies." *Nature*, 531(7595): 496–499.

Galbiati, Roberto, Emeric Henry, and Nicolas Jacquemet. 2017. "Spillovers, Persistence and Learning: Institutions and the Dynamics of Cooperation." *CEPR Discussion Papers*, DP12128.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives*, 25(4): 191–210.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

Grimm, Veronika, and Friederike Mengel. 2012. "An experiment on learning in a multiple games environment." *Journal of Economic Theory*, 147(6): 2220–2259.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2008. "Alfred Marshall: Lecture Social Capital As Good Culture." *Journal of the European Economic Association*, 6(2-3): 295–320.

Gürerk, Ozgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. "The competitive advantage of sanctioning institutions." *Science*, 312(5770): 108–111.

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Richard McElreath, and Herbert Gintis. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review: Papers & Proceedings*, 91(2): 73–78.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science*, 319(5868): 1362–1367.

Jackson, Matthew, and Daron Acemoglu. 2017. "Social Norms and the Enforcement of Laws." *Journal of the European Economic Association*, 15(2): 245–295.

Jehiel, Philippe. 2005. "Analogy-based expectation equilibrium." *Journal of Economic Theory*, 123(2): 81–104.

Kamei, Kenju. 2016. "Democracy and resilient pro-social behavioral change: an experimental study." *Social Choice and Welfare*, 47(2): 359–378.

Keser, Claudia, Andreas Markstädter, and Martin Schmidt. 2017. "Mandatory minimum contributions, heterogeneous endowments and voluntary public-good provision." *Games and Economic Behavior*, 101: 291 – 310.

Kimbrough, Erik O., and Alexander Vostroknutov. 2016. "Norms Make Preferences Social." *Journal of the European Economic Association*, 14(3): 608–638.

Kimbrough, Erik O., and Alexander Vostroknutov. 2017. "A Portable Method of Eliciting Respect for Social Norms." *Working Paper*.

Knez, Marc, and Colin Camerer. 2000. "Increasing Cooperation in Prisoner's Dilemmas by Establishing a Precedent of Efficiency in Coordination Games." *Organizational behavior and human decision processes*, 82(2): 194–216.

Kosfeld, Michael, Akira Okada, and Arno Riedl. 2009. "Institution Formation in Public Goods Games." *American Economic Review*, 99(4): 1335–1355.

Krupka, Erin L., and Roberto A. Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11(3): 495–524.

Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. 2019. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the National Academy of Sciences*, 116(10): 4156–4165.

López-Pérez, Raúl. 2008. "Aversion to norm-breaking: A model." *Games and Economic Behavior*, 64(1): 237–267.

McCarter, Matthew W., Anya Samek, and Roman M. Sheremeta. 2014. "Divided Loyalists or Conditional Cooperators? Creating Consensus About Cooperation in Multiple Simultaneous Social Dilemmas." *Group & Organization Management*, 39(6): 744–771.

Nakashima, Nathaniel A., Eliran Halali, and Nir Halevy. 2017. "Third parties promote cooperative norms in repeated interactions." *Journal of Experimental Social Psychology*, 68: 212–223.

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

Ostrom, Elinor. 1990. *Governing the Commons - The Evolution of Institutions for Collective Action.* New York:Cambridge University Press.

Ostrom, Elinor, James Walker, and R Gardner. 1992. "Covenants With and Without a Sword: Self-Governance is Possible." *The American Political Science Review*, 86(2): 404–417.

Peysakhovich, Alexander, and David G. Rand. 2016. "Habits of virtue: creating norms of cooperation and defection in the laboratory." *Management Science*, 62(3): 631–647.

Peysakhovich, Alexander, and David G. Rand. 2017. "In-group favoritism caused by Pokémon Go and the use of machine learning for principled investigation of potential moderators." *Working Paper*.

Peysakhovich, Alexander, and Jeffrey Naecker. 2017. "Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity." *Journal of Economic Behavior & Organization*, 133: 373–384.

Posner, Richard A. 1997. "Social norms and the law: An economic approach." *The American Economic Review*, 87(2): 365–369.

Savikhin, Anya C., and Roman M. Sheremeta. 2013. "Simultaneous Decision-Making in Competitive and Cooperative Environments." *Economic Inquiry*, 51(2): 1311–1323.

Schlag, Karl H., and Joël J. van der Weele. 2015. "A method to elicit beliefs as most likely intervals." *Judgment and Decision Making*, 10(5): 456–468.

Selten, Reinhard. 1967. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments." In *Beiträge zur experimentellen Wirtschaftsforschung*. Vol. I, , ed. Heinz Sauermann, 136–168. Tübingen:JCB Mohr (Paul Siebeck).

Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97(3): 999–1012.

Stagnaro, Michael N., Antonio A. Arechar, and David G. Rand. 2017. "From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement." *Cognition*, 167: 212–254.

Sunstein, Cass R. 1996. "On the Expressive Function of Law." *University of Pennsylvania Law Review*, 144(5): 2021–2053.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher. 2010. "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *Review of Economic Studies*, 77(4): 1540–1566.

Tabellini, Guido. 2008. "The Scope of Cooperation: Values and Incentives." *The Quarterly Journal of Economics*, 123(3): 905–950.

Tabellini, Guido. 2010. "Culture and Institutions: Economic Development in the Regions of Europe." *Journal of the European Economic Association*, 8(4): 677–716.

Tyran, Jean-Robert, and Lars P. Feld. 2006. "Achieving Compliance when Legal Sanctions are Non-deterrent." *Scandinavian Journal of Economics*, 108(1): 135–156.

Vischer, Thomas, Thomas Dohmen, Armin Falk, David Huffman, Jürgen Schupp, Uwe Sunde, and Gert G. Wagner. 2013. "Validating an ultra-short survey measure of patience." *Economics Letters*, 120(2): 142–145.