LEMMA – Document de travail

DT 2026-05

Inefficiency is worse for cooperation than corruption: Experimental evidence

Marina Povitkina

University of Gothenburg

Andrea F.M. Martinangeli

Université Paris Panthéon-Assas, LEMMA

Sverker C. Jagers

University of Oslo

LEMMA EA 4442 - 4 rue Blaise Desgoffe 75006 Paris

Inefficiency is worse for cooperation than corruption: Experimental evidence

Marina Povitkina[1,3], Andrea F.M. Martinangeli[2,3], and Sverker C. Jagers[1,3]

[1]University of Gothenburg
[2]Université Paris 2 Panthéon-Assas
[3]University of Oslo

Abstract

We test whether corruption and institutional inefficiency affect cooperation differently. Both reduce returns to collective action, but corruption entails private appropriation, whereas inefficiency dissipates resources without benefiting an identifiable actor. We test this distinction in a large-scale, lab-like public-good experiment in Germany and Romania, in which we randomly vary the mechanism through which group resources fail to be redistributed. Contrary to our expectation, participants contribute less when resources may be lost through inefficiency than when an administrator can embezzle them. This pattern is clear and robust in Germany and similar, though smaller and less precisely estimated, in Romania. Institutional resource loss may undermine cooperation at least as strongly as corruption. The findings highlight the behavioral consequences of failures of institutional, not only violations of institutional integrity.

Acknowledgements The authors are grateful to Oleksii Polovnikov for research assistance. We thank Federica Braccioli, Xavier Gassmann, Roberto Hernan-Gonzalez, Niklas Harring, Steffen Kallbekken, Patrick Michaelsen, Marina Nistotskaya, Jonas Nässen, Eli Spiegelman, Angela Sutan, Nina Westberg, and the researchers at the Quality of Government Institute for valuable comments and suggestions. The project received funding from the Swedish Research Council, grant #2019-02680 and Norwegian Research Council, grant #354071

1 Introduction

There is a general belief that in the presence of dysfunctional political institutions, for example suffering from corruption and inefficiencies, individuals stop trusting each other (Rothstein and Eek, 2009), stop trusting the government (Harring, 2016), and tend to focus on their own benefits rather than collective well-being (Persson et al., 2013a). Indeed, when people's contributions to the system (e.g., taxes that should finance public service provision) get stolen, they have few reasons to believe that others would contribute and little motivation to contribute themselves: On the one hand, the unethical nature of theft may harm individuals' morale; on the other, the subtraction of resources from the public coffers is analogous to an inefficiency in which the money is lost in a system that does not deliver. In this paper, we aim to disentangle these two motivations by asking whether corruption has a smaller, equal, or larger detrimental effect on cooperation and people's willingness to contribute to collectively beneficial outcomes than systemic inefficiencies.

Both corruption and institutional inefficiency are broad phenomena that take different forms and arise from different causes. Corruption can involve bribery, kickbacks, stealing of public funds, patronage, among other practices. Inefficiency may stem from administrative incapacity, organizational shortcomings, poor coordination, or waste. In this study, we focus on one aspect of corruption: embezzlement of public funds; and one specific comparable dimension of inefficiency: institutional resource loss, whereby resources devoted to a collective purpose fail to generate corresponding collective value. We thus ask whether private appropriation of public funds affects people's willingness to contribute differently than the loss of these public funds due to inefficiencies.

This question is non-trivial. In purely economic terms, corruption and inefficiency are similar because in both cases individuals experience an economic loss in the form of lower returns to their contributions to the public good. Because people receive a lower return in public-good value than they would otherwise expect, both institutional shortcomings could have similar effects on cooperative behaviour. At the same time, we argue that perceptions of the two shortcomings may differ and thereby produce different behavioural reactions. In the case of corruption, people face a situation in which their contributions are deliberately misappropriated by an entrusted actor for private benefit. In the case of inefficiency, resources instead fail to generate collective value without being privately appropriated. Thus, while corruption signals a failure of institutional integrity, inefficiency may instead signal a failure of institutional competence or capacity. We expect corrupt practices to be more detrimental than institutional resource loss for people's willingness to cooperate.

After theorizing the potentially different effects of corruption and inefficiency, we test our expectations using the same large-scale, lab-like experiment in two different political contexts: In Germany, where corruption and inefficiencies are not expected to be the norm, and Romania, where corruption and inefficiencies are likely to be considered common. We design a novel behavioural game modelling institutional resource loss due to inefficiencies and modify the recently developed embezzlement game (Martinangeli

et al., 2024) to enable comparison. Participants within each country are assigned to experimental conditions that vary the mechanism through which resources fail to be redistributed. Random assignment allows us to identify the causal effects of these experimental conditions on cooperation within each national sample. Implementing the same experiment in two countries also allows us to examine whether the treatment effects differ between the different institutional contexts. While any such difference cannot by itself identify the effect of national institutional quality, it would provide valuable information on whether people living in different institutional contexts react to experimentally induced institutional dysfunctionalities.

Contrary to our theoretical expectation, cooperation is more adversely affected by institutional resource loss than by the possibility of embezzlement. The participants contribute less when public resources may be lost through systemic inefficiency than when an Administrator can appropriate them. This pattern is clear and robust in the German sample and directionally similar, although smaller and less precisely estimated, in Romania. The findings suggest that impersonal institutional waste can undermine cooperation at least as strongly as intentional private appropriation.

We contribute to the literature in three important ways. First, we theorize about the different effects of inefficiencies and corruption on people's willingness to cooperate. Second, we provide empirical evidence for how these two dysfunctionalities shape cooperation using a unique experimental data collection. Related experimental evidence indicates that corruption-related costs need not discourage investment more than comparable legitimate costs (Banerjee et al., 2022). We examine whether an analogous pattern arises in a cooperative public-good setting. Third, we suggest a novel experimental design modelling institutional corruption and inefficiencies available for replication by scholars researching the effect of institutional quality on people's cooperation.

2 Previous research

Growing evidence shows that people trade off prosocial and self-regarding behaviour differently across institutional settings of different quality (e.g., Herrmann et al., 2008; Gächter and Herrmann, 2009; Engl et al., 2021). Cross-societal research documents substantial variation in cooperative and sanctioning behaviour across societies (e.g., Steinmo and D'Attoma, 2022; Rothstein and Eek, 2009), suggesting that cooperation is not simply a function of individual preferences but is also shaped by the institutional and normative environments in which social interactions take place (e.g., Davidovic et al., 2020). Such studies provide broad comparative coverage and can establish whether behaviours and attitudes covary with institutional conditions across societies. Their causal interpretation is nevertheless limited because institutional quality is correlated with historical, cultural, economic, and political characteristics that may also affect individual behaviour.

The distinction between cross-country and experimental research is separate from the distinction between correlational and causal research. Experiments may be conducted within one country or implemented comparatively across several countries. Their

particular advantage is that random assignment permits causal inference about the experimentally manipulated features. Comparative experiments can additionally establish whether those treatment effects differ across national samples, although they do not, without further assumptions, identify which country-level characteristics cause such differences.

One important factor influencing people's cooperative choices found in both cross-sectional and experimental research is corruption: its effect on trust among individuals (e.g., Rothstein and Eek, 2009; Dinesen, 2012; Martinangeli et al., 2024) and on trust in the institutions individuals interact with (Rothstein, 2011) ultimately erodes cooperation (e.g., Campos-Vazquez and Mejia, 2016; Muthukrishna et al., 2017). Embezzlement, or misuse of public funds, the form of corruption this article focuses on, while under-investigated, has also been found to reduce trust and cooperation (Attanasi et al., 2019; Martinangeli et al., 2024; Spadaro et al., 2023).

Another factor influencing cooperative choices is institutional inefficiency, although much less researched than corruption. Experimental evidence suggests that individuals respond negatively when resources devoted to collective purposes are partly lost in the process of redistribution. Tepe et al. (2021), for example, experimentally manipulate efficiency loss during redistribution and find that greater efficiency loss reduces support for redistribution. Their design is particularly relevant here because it isolates efficiency loss itself rather than tying the loss to a specific form of institutional wrongdoing. A related literature on public-sector performance shows that citizens' experiences and evaluations of government performance are associated with institutional trust, although the causal relationship between performance and trust is complex and potentially reciprocal (Van de Walle and Bouckaert, 2003; Kampen et al., 2006). Similarly, uncertainty about the outcome—about others' contributions, payoffs, or the likelihood that a collective goal will be reached—has been found to reduce cooperation (Van Dijk et al., 2004), especially if individuals believe their contribution to be critical to attaining the collectively beneficial outcome (Chen et al., 1996).

The closest experimental comparison is Banerjee et al. (2022), who distinguish the effect of corruption-related costs from that of legitimate costs in an investment experiment in Kenya. Both types of costs discouraged investment, but corruption did not impose an additional deterrent effect beyond the material cost itself. Their study therefore shows that the moral character of a loss need not make it more behaviourally consequential than an otherwise comparable cost. We extend this comparison from investment to cooperation in a public-good setting, contrasting intentional private appropriation with impersonal institutional resource loss under comparable direct payoff consequences.

In all, experimental research manipulating specific dimensions of institutional quality, especially institutional inefficiency, remains very limited. Existing observational and cross-country research provides important evidence about associations between institutional conditions and individual behaviour, but is less well suited to isolating the effects of particular institutional features. We address this research gap in the literature by both randomly manipulating institutional quality or the way we operationalize it - the rules enabling or hindering resource loss - and by implementing the same study in dif-

ferent national settings. Our design therefore provides causal estimates of the effects of the source of public money loss within each sample, while the comparison between contexts of different institutional quality levels examines cross-sample heterogeneity in those effects.

3 Theory

Corruption is broadly defined as the misuse of entrusted power for private gain (Transparency International, 2025). It encompasses a variety of practices, including bribery, embezzlement, and more systemic forms such as patronage or state capture. By contrast, institutional efficiency refers to the relationship between inputs and outputs in public administration. Governments are considered efficient when they convert resources into public goods effectively and with minimal waste (Tanzi and Davoodi, 1998; Rajkumar and Swaroop, 2008). Institutional inefficiency can consequently take several forms and might come from different causes. It may result from administrative incapacity, organizational shortcomings, poor coordination, faulty procedures, or other forms of waste. It is beyond the scope of one study to capture this entire conceptual domain. We therefore isolate one consequence common to many such failures: institutional resource loss, in which resources devoted to a collective purpose fail to generate corresponding collective value. Efficiency and effectiveness can in this sense be analytically distinguished from other dimensions of institutional quality, such as the impartial exercise of public authority (Rothstein and Teorell, 2008). We argue that while both corruption and inefficiency can result in lower returns from public-good contributions, they differ in their underlying mechanisms and normative implications and may therefore affect cooperative behaviour differently.

In purely economic terms, the two phenomena create similar incentives - individuals contributing to a public good receive less in return than they reasonably expect. Thus, whether public resources are stolen by corrupt officials or lost due to inefficiencies, the material outcome for citizens is a reduction in the effectiveness of their contributions. From a rational choice perspective, this reduction should weaken incentives to contribute to the public good because the expected payoff from cooperation decreases regardless of the underlying cause of the economic loss. However, we argue that there are several reasons to believe corruption and inefficiency should influence behaviour differently, even if the experienced economic losses are similar.

First, an important difference between corruption and inefficiency concerns the attribution of responsibility. Corruption involves intentional wrongdoing, often by identifiable actors who deliberately misuse public resources for private gain. Because the behaviour is purposeful, citizens can attribute blame to specific individuals or groups within the political and administrative system. Inefficiency, by contrast, may be perceived as the result of systemic complexity, bureaucratic inertia, or organizational shortcomings rather than deliberate exploitation. Responsibility is therefore more diffuse and difficult to assign to a particular actor. Intentionality, however, is not the only distinction involved. Corruption, by definition, entails a violation of entrusted authority: resources

that should be administered according to institutional rules are instead appropriated for private gain. Inefficiency may also undermine institutional legitimacy, particularly when poor performance is persistent, but it does not necessarily involve such a violation. The distinction can therefore also be understood as one between failures primarily signalling compromised institutional integrity and failures primarily signalling deficient institutional competence or capacity (Rothstein and Teorell, 2008; Van de Walle and Bouckaert, 2003).

Research on blame attribution suggests that people respond more negatively to losses caused by intentional wrongdoing than to similar losses resulting from structural or accidental processes (Shaver, 1985; Alicke, 2000). Intentional violations can therefore be expected to generate stronger blame and moral condemnation, whereas losses attributed to impersonal or structural failures may generate different and potentially weaker moral reactions. For this reason, corruption may provoke stronger behavioural responses, including withdrawal from cooperative arrangements that rely on trust in others' good intentions. At the same time, the ability to attribute responsibility may also influence perceptions of accountability and control (Braccioli et al., 2025). When corruption is present, individuals may believe that the problem is rooted in actors who could, at least in principle, be removed or punished. In contrast, inefficiency often reflects structural characteristics of institutions that appear more difficult to reform.

The second important distinction concerns the types of norms that corruption and inefficiency violate. Political philosophers and social theorists distinguish between moral norms and social norms (Brennan et al., 2013). Moral norms refer to fundamental principles about right and wrong that individuals perceive as intrinsically binding. Violating a moral norm typically triggers strong moral condemnation and social sanction. Social norms, by contrast, describe commonly accepted patterns of behaviour within a society or organization. They regulate behaviour but are often more flexible and context dependent.

Corruption directly violates a moral norm (Banerjee, 2016). Public officials are entrusted with the authority to act in the public's interest, and the misuse of this authority for private gain constitutes a clear breach of that moral entrustment. Because moral norms have an "either–or" character, their violation tends to generate strong and lasting reputational consequences. After experiencing corruption, trust in the perpetrator is fundamentally damaged, and the legitimacy of the institution they represent may also be called into question. Corruption signals that actors entrusted with authority are willing to violate fundamental principles for personal benefit, which can have detrimental effects on people's cooperative behaviour, because it undermines citizens' broader normative expectations about behaviour within the system (Rothstein, 2013).

Efficiency, in contrast, concerns institutional performance: the capacity of institutions to transform available resources into expected outputs. Poor performance may reduce citizens' satisfaction with public institutions and undermine institutional trust (Van de Walle and Bouckaert, 2003; Kampen et al., 2006), but it does not necessarily imply that an entrusted actor has deliberately violated institutional rules for private benefit. We therefore expect the normative signal generated by inefficiency to differ

from that generated by corruption.

In addition, when individuals encounter corruption, they may infer that the broader social environment is characterized by opportunistic behaviour and weak moral constraints. Such inferences can erode both institutional trust and, in turn, interpersonal trust (Rothstein, 2013; Rothstein and Eek, 2009; Spadaro et al., 2023; Martinangeli et al., 2024), through beliefs about the trustworthiness of others (Banerjee, 2016). If citizens believe that others are likely to exploit the system for personal gain, they may become less willing to contribute to collective endeavours themselves.

Recent research demonstrates that exposure to corrupt environments can have such spillover effects. Institutions that leave room for corruption can reduce individuals' willingness to trust others in unrelated social interactions (Martinangeli et al., 2024). In this sense, corruption may have a demoralizing effect that extends beyond the institutional setting in which it occurs. Instead of merely reducing the material returns to cooperation, it also weakens the normative expectations that sustain cooperative behaviour in the first place.

Inefficiency, by contrast, does not necessarily convey the same social signal. When resources are lost due to administrative or organizational shortcomings, the failure may instead signal deficient institutional capacity or performance. Such performance failures can reduce satisfaction with public institutions and may contribute to lower institutional trust (Van de Walle and Bouckaert, 2003; Kampen et al., 2006). They do not, however, necessarily imply that other individuals are acting opportunistically. Consequently, we expect their spillover effects on interpersonal trust, and thereby cooperative behaviour, to be weaker than those generated by corruption.

Another perspective through which the two phenomena differ is the distinction between *procedural* and *distributive* fairness. Distributive fairness concerns the perceived fairness of outcomes, while procedural fairness relates to the perceived fairness of the processes via which outcomes are generated (Grimes, 2006; Akbaş et al., 2019; Spadaro et al., 2023).

Both corruption and inefficiency can contribute to distributive unfairness, and persistent inefficiency may also undermine perceptions of institutional competence and ultimately institutional legitimacy (Van de Walle and Bouckaert, 2003; Kampen et al., 2006). The distinction is therefore not that inefficiency affects outcomes whereas corruption alone affects procedures. Rather, corruption contains an additional and unambiguous violation of entrusted authority: an actor responsible for administering collective resources deliberately redirects them toward private benefit.

Perceptions of procedural unfairness are known to have particularly strong effects on cooperation and compliance (Sun et al., 2022). When individuals believe that decision-making processes are unfair or manipulated, they may withdraw their support for collective institutions altogether. Inefficiency can likewise undermine confidence in institutions, but it does not necessarily signal that entrusted actors have violated institutional rules for private benefit. In this sense, corruption represents both a performance failure and an integrity violation, whereas inefficiency can represent a performance failure without necessarily involving a violation of integrity.

Taken together, these theoretical considerations suggest that corruption should have stronger negative effects on cooperative behaviour than inefficiency. Although both phenomena reduce the returns from contributing to the public good, corruption additionally involves intentional private appropriation and an explicit violation of entrusted authority. Institutional resource loss may undermine confidence in institutional competence and performance, but does not necessarily convey the same signal of opportunistic behaviour. For these reasons, we expect corruption to produce a stronger decline in cooperation. We therefore hypothesize that:

H1: Institutional corruption damages cooperation among individuals more than institutional resource loss resulting from inefficiency.

While the mechanisms described above should hold in general, their strength may depend on the broader institutional context in which individuals operate. Citizens develop expectations about how institutions typically function based on their everyday experiences within a particular political system. These experiences shape beliefs about the prevalence of corruption, the effectiveness of public administration, and the trustworthiness of others (Rothstein, 2013).

In contexts where corruption is rare and institutions function relatively efficiently, exposure to corruption may represent a particularly strong violation of existing expectations. Conversely, where corruption and administrative shortcomings are more prevalent, individuals may be more likely to anticipate such institutional failures and incorporate them into their expectations about institutional functioning.

This reasoning is consistent with research suggesting that institutional environments can shape behavioural expectations and norms (Fisman and Miguel, 2007; Barr and Serra, 2010; Persson et al., 2013b). We therefore expect the behavioural consequences of corruption and inefficiency to vary across institutional contexts.

H2: The negative effect of corruption on cooperation (relative to inefficiency) will be stronger in high institutional quality settings than in low institutional quality settings.

Hypothesis 2 calls for investigating two national contexts, one characterised by higher institutional quality and one by lower institutional quality. We caution, however, that while a potential difference between the two contexts would be consistent with the argument underlying H2, such effect heterogeneity would not by itself establish which contextual characteristic generated that difference.

4 Research design

4.1 Experimental design

We study cooperation using a modified public-good game based on the embezzlement game introduced by Martinangeli et al. (2024). Participants were assigned to three-person groups comprising one Administrator and two Group Members. Each participant received an individual endowment E, denominated in the local currency, and indepen-

dently decided how much of this endowment to contribute to a group project.[1] The Administrator also made a contribution decision.

Let c_i denote participant i's contribution and let

$$C = \sum_{j \in G} c_j$$

denote total group contributions. In the absence of any deduction, contributions were doubled and divided equally among the three group members. Thus, participant i's payoff in the standard version of the game was

$$\pi_i = E - c_i + \frac{2}{3}C. \tag{1}$$

Because each monetary unit contributed returned 2/3 of a unit to each participant, contributing increased total group earnings but reduced the contributor's private payoff, holding the contributions of the other group members fixed.

The modified game placed the group contributions under the administration of one member of the group. Depending on the experimental condition, the Administrator was either required to redistribute the entire group budget, thus reproducing the standard game (Condition *Baseline*) or was permitted to retain any part of it, enabling embezzlement of public funds (Condition *Unconstrained*). In two further conditions, the software could instead prevent part of the budget from being redistributed, and exogenously varied whether the non-redistributed amount accrued to the administrator (Condition *Involuntary*) or "was lost" (Condition *Inefficiency*). The experimental design therefore compares willingness to contribute to the public good when the institutions perfectly prevent embezzlement (*Baseline*) with willingness to contribute when Administrators are free to embezzle in *Unconstrained*. Further, it disentangles the effects of simply receiving a lower payoff from the Administrators' unethical action by introducing Conditions *Involuntary* and *Inefficiency*. In doing so, it exploits the fact that from a non-Administrator Group Member's perspective, encountering embezzlement and inefficiencies brings identical monetary consequences. The two differ only in terms of their impact on total group earnings and in the intentionality of the administrator's embezzlement. These implications are discussed in the next paragraph.

Participants progressed through the experiment in the following steps:

1. All participants received a common explanation of the standard public-good game, including the contribution and payoff rules.

2. All participants received an explanation of the Administrator's role in collecting and redistributing group contributions.

3. Each participant was assigned to one of the four experimental conditions and received only the instructions corresponding to that condition. Participants could

[1] The endowment E given to each participant amounted to 2 € and 10 lei, approximately equivalent at the time of the experiment, respectively in the German and Romanian implementations.

rewatch the instructional videos or inspect still images reproducing the instructions at their own pace.

4. Participants completed comprehension questions concerning the calculation of payoffs in the public-good game.

5. All participants, including those assigned the Administrator role, made their contribution decision without observing the contributions of the other group members.

6. The redistribution screen was displayed only to Administrators. Administrators indicated the percentage of group contributions that they would declare for redistribution. Whether they could declare less than 100%, and what happened to any amount not redistributed, depended on the experimental condition.

7. Participants completed incentivised belief-elicitation questions. Non-Administrator Group Members stated their beliefs about the contribution of the other Group Member and the percentage of group contributions declared for redistribution by the Administrator.

The public-good game was a one-shot interaction among anonymous participants. It was implemented asynchronously: participants made their decisions without observing the contemporaneous decisions of the other group members and were subsequently matched with participants assigned to the same experimental condition. The matching procedure is described in Appendix H.

The Administrator and amounts not redistributed. Let R, with $0 \leq R \leq C$, denote the amount of the group budget not made available for redistribution. A non-Administrator Group Member's payoff was therefore

$$\pi_i = E - c_i + \frac{2}{3}(C - R), \qquad i \neq k, \tag{2}$$

where k denotes the Administrator.

For a given value of R, the direct payoff consequence for a non-Administrator was the same whether R was retained by the Administrator or lost through the software.[2] The Administrator's payoff depended additionally on the destination of R:

$$\pi_k = E - c_k + \frac{2}{3}(C - R) + \delta_z R, \tag{3}$$

where z denotes the experimental condition and

$$\delta_z = \begin{cases} 1, & \text{if the amount } R \text{ accrued to the Administrator,} \\ 0, & \text{if } R = 0 \text{ or the amount } R \text{ was lost.} \end{cases}$$

[2]Shown formally in Appendix A.

Consequently, the conditions could impose the same direct reduction in redistribution on non-Administrators while differing in the origin and destination of the non-redistributed amount. When R accrued to the Administrator, it remained part of total group earnings; when it disappeared through the software, it constituted a loss for the group as a whole. In particular, group earnings remain higher when R is embezzled, because the amount stays in the Administrator's purse, than when it is lost through inefficiencies.

We test *Hypothesis 1* by investigating the behavioural consequences of embezzlement and inefficiency that have identical payoffs. We do so by disentangling embezzlement intentionality (*Unconstrained* vs *Involuntary*) from their relative impact on total group earnings (*Involuntary* vs *Inefficiency*).

Appendix A provides the formal comparison of aggregate payoffs. Appendix B shows that introducing R, irrespective of its nature, does not change the Nash-equilibrium prediction of zero contributions.

Experimental conditions. Participants were assigned to one of four conditions. Table 1 summarizes the Administrator's permitted action, the role of the software, the destination of any amount not redistributed, and the information provided before the contribution decision.

In the *Baseline* condition, the Administrator had neither the opportunity to retain group resources nor the ability to reduce redistribution. This condition therefore implemented full redistribution of the group budget. Theoretically, the Baseline condition represents an efficient and non-corrupt institutional benchmark within the experiment. Both institutional integrity and institutional performance are intact: the Administrator cannot appropriate collective resources, and the full group budget is converted into redistribution. It is not intended to represent an ideal institution more generally, but rather a benchmark in which neither of the institutional dysfunctionalities examined in the experiment is present.

In the *Unconstrained* condition, the Administrator could declare less than the full group budget for redistribution and retain the undeclared amount. The amount R was therefore determined by the Administrator's decision. The amount retained was not disclosed to the two non-Administrator Group Members. The *Unconstrained* condition therefore represents institutions that allow for a possibility of embezzlement.

In the *Involuntary* condition, the Administrator was required to declare the full group budget, but the software determined how much was actually redistributed. The amount not redistributed accrued to the Administrator, although its size did not result from the Administrator's decision. This condition was designed to preserve the private allocation of R to the Administrator while removing the Administrator's discretion over that allocation.

In the *Inefficiency* condition, the Administrator was again required to declare the full group budget. The software, however, removed part of the budget before redistribution. The removed amount did not accrue to the Administrator or to another participant. This condition therefore implemented software-generated (institutional) resource loss rather than private appropriation.

Table 1: Experimental conditions and information provided to participants

Condition	Administrator's permitted action	Software rule	Destination of R	Information provided before the decision
Baseline	The Administrator was required to declare the entire group budget for redistribution.	The software did not remove any part of the budget; $R = 0$.	No amount was withheld or lost.	Participants were informed that the Administrator was required to redistribute the entire group budget.
Unconstrained	The Administrator could declare any percentage of the group budget for redistribution and retain the remainder.	The software did not determine R; the amount depended on the Administrator's declaration.	The Administrator.	Participants were told that, in a previous similar study, Administrators had redistributed 70% of group contributions and retained 30% on average.
Involuntary	The Administrator was required to declare the entire group budget for redistribution.	The software determined the amount actually redistributed, leaving the remaining amount to the Administrator.	The Administrator.	Participants were told that the software would redistribute 70% of group contributions and leave 30% to the Administrator on average.
Inefficiency	The Administrator was required to declare the entire group budget for redistribution.	The software removed part of the budget before redistribution, independently of the Administrator's decision.	The amount disappeared and did not accrue to the Administrator or another group member.	Participants were told that the software would redistribute 70% of group contributions and that 30% would be lost on average.

The software-generated loss should not be interpreted as a literal representation of how institutional inefficiency in general occurs in real-world settings. Real-world inefficiencies may be systematic and become visible through, for example, administrative failures, cost overruns, delayed services, or low-quality public goods. Our treatment abstracts from these specific manifestations and instead isolates a consequence they can share: resources devoted to a collective purpose fail to generate corresponding collective value. The *Inefficiency* condition should therefore be understood as an experimental representation of institutional resource loss, rather than as a comprehensive simulation of institutional inefficiency.

Information about expected redistribution. Because participants made their contribution decisions without observing the subsequent realization of R, we provided a common numerical benchmark for expected redistribution in the three conditions in which a reduction was possible. The benchmark was based on the redistribution decisions observed in Martinangeli et al. (2024).

In the Unconstrained condition, participants were told:

> In a previous similar study, the Administrators distributed, on average, 70% of the group contributions and kept 30% for themselves.

In the Involuntary condition, participants were told that the software would, on average, distribute 70% of group contributions and leave 30% to the Administrator. In the Inefficiency condition, they were told that the software would, on average, distribute 70% of group contributions and that 30% would be lost. Thus, the numerical benchmark was held constant across the three conditions, whereas the stated source and destination of the reduction varied.

Providing the same numerical benchmark does not imply that realized redistribution was identical across conditions. In the Unconstrained condition, the realized value of R depended on the Administrator's declaration. In the Involuntary and Inefficiency conditions, it was determined by the software, which randomly drew and implemented values from the full distribution of embezzlement rates observed in Martinangeli et al. (2024).

Online implementation and materials. The experiment was programmed in oTree (Chen et al., 2016) and fielded in German and Romanian languages (see next section for the case selection rationale). Instructions were delivered through animated videos illustrating the public good game, the Administrator's role, and the participant's assigned condition. Before making their decisions, participants could rewatch the videos or inspect still images reproducing the same instructions at their own pace. They subsequently completed comprehension questions concerning the calculation of earnings.

After having viewed the explanation of the general public good game, the participants were confronted with a battery of 6 control questions (See Appendix F) intended to gauge their understanding of the simple public good game mechanics and payoff computations. These questions were simplified to adapt them to the online setting. We introduced

a counter for the number of times a participant submitted the control questions form containing at least an error. After the 10th incorrect attempt, the participant was screened out and redirected out of the survey. On average, the participants submitted the form twice with at least one mistake.

Appendix F reproduces the complete English translations of the participant-visible narration and on-screen text for the common instructions and each of the four conditions. It also reproduces the contribution, redistribution, correction, and belief-elicitation screens. Links to the videos are provided as supplementary materials rather than as substitutes for the written instructions.

The survey also included the elicitation of individual-level social trust using the trust game (Berg et al., 1995), social norms of cooperativeness, trust, and trustworthiness using a variant of the opinion matching method (Bicchieri et al., 2021), and risk preferences using the bomb task by Crosetto and Filippin (2013). These elicitations all took place after the conclusion of the public good game (the main task of interest in this study) and were presented in random order. Further details about their implementation are presented in Appendix D.

4.2 Country contexts and data collection

When evaluating the effects of different aspects of institutional quality—in our case, the possibility of embezzlement and institutional resource loss—it is important to recognize that participants may enter the experiment with different prior experiences of institutional functioning.

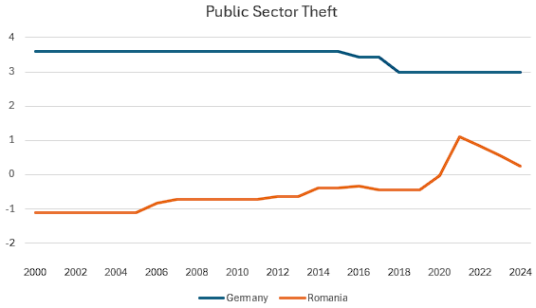
We implemented the same experiment in Germany and Romania to examine whether the experimental treatment effects are similar across two national settings with different institutional histories. Germany and Romania differ in numerous historical, economic, cultural, and political respects in addition to the institutional dimensions considered here. Country comparative results should thus be interpreted with caution.

The selection of Germany and Romania was motivated by their differences in the institutional phenomena studied. Germany is characterised by comparatively low levels of public-sector embezzlement and relatively low levels of institutional resource loss, whereas Romania has historically recorded higher levels of public-sector embezzlement and higher levels of institutional resource loss. Figure 1 presents indicators commonly used to measure these two aspects of institutional quality to document this contextual contrast.

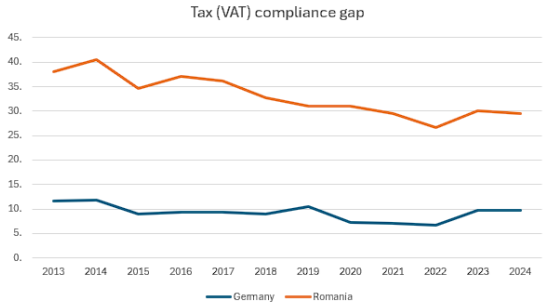
We use the Varieties of Democracy Public Sector Theft Index, which captures the extent to which public-sector employees grant favours in exchange for material goods and how often they steal, embezzle, or misappropriate public funds or other state resources for personal or family use.[3] A higher score on the index implies a public administration

[3]There are various ways to measure corruption at the national level. The most popular are expert-based measures, such as Worldwide Governance Indicators by the World Bank and V-Dem, and citizen-based measures, such as Corruption Perception Index (CPI) by Transparency International and European Quality of Government Index (EQI) by the Quality of Government Institute. In comparing countries, we chose the measure closest to our experimental treatment, capturing embezzlement in the public ad-

that is relatively free from embezzlement practices. An indicator of institutional resource loss comparable across national contexts is much harder to find. We turn to the European Union's VAT compliance gap, which captures the difference between expected VAT revenues under full compliance and revenues actually collected (European Commission, 2025). The VAT compliance gap is not a direct indicator of institutional inefficiency. It may reflect several factors, including taxpayer behaviour, enforcement, administrative capacity, and features of the tax system. Nevertheless, it illustrates the extent to which potential public revenues are successfully collected, and thereby the extent of potential resource loss, making it possible to compare countries.



Source: Varieties of Democracy (V-Dem) Institute (2025)



Source: European Commission (2025)

Figure 1: Public sector theft 2000-2024 (left) and Tax compliance gap 2013-2024 (right).

The figure illustrates that Germany has had a public administration nearly free from embezzlement practices. In contrast, Romania has scored much lower on the index, with an improvement only in recent years. Germany's VAT compliance gap is currently approximately 9%, close to the European average, whereas Romania's is the highest in the EU and reaches approximately 30%. The indicators thus document substantial differences between the two country contexts in public-sector embezzlement and in the extent to which potential tax revenues are successfully collected. The latter should not, however, be interpreted as a direct measure of institutional inefficiency, since the VAT compliance gap reflects both institutional and behavioural factors. However, it still illustrates the differences in the potential institutional resource loss in the two countries.

These contextual differences motivate the expectation that participants in the two samples may enter the experiment with different beliefs about institutional functioning. Under the argument developed in Hypothesis 2, participants in the two samples may therefore respond differently to the experimental conditions if those conditions depart

ministration from an expert-based survey (V-Dem). When comparing countries on the citizen-based indicators of corruption, which measure the broader concept of corruption, the picture is similar: Germany scores around 80 out of 100 on CPI, while Romania scores 45, with higher scores corresponding to lower levels of corruption. Similarly, Germany scores around 1 on -3 to 3 scale on EQI, while Romania scores around -1.5, with higher scores corresponding to higher levels of institutional quality.

to different degrees from their prior expectations. Differences between the estimated treatment effects while potentially reflecting differences between the two settings other than just the quality of their institutions, would lend credibility to these arguments.

We collected data using online surveys coded with oTree (Chen et al., 2016). Such implementation made a lab-like experiment possible on a large sample representative along the age, gender, region of residence, and income dimensions. Previous studies comparing laboratory experiments to such lab-like online survey experiments find few differences in the results, implying that the results from both are comparable (Buso et al., 2021; Clifford and Jerit, 2014).

In both Romania and Germany, we used the Respondi/Bilendi survey company for sampling and survey data collection. In Germany, we collected data from 4,812 respondents. In Romania, our sample consists of 3,007 respondents. The data collection took place in 2023–2024.

5 Results

First, we compare how individuals' beliefs about the contributions of others and the administrator's honesty differ between the experimental conditions. We can thus gain a sense of whether the experimental conditions had the desired effect on individuals' expectations about the behaviour of others, that is, whether institutions differing in their control over administrators' behaviours change beliefs about the behaviour of others. Due to the nature of this comparison, we set Baseline treatment condition as a reference category. All reported OLS regressions include a common set of controls: age categories, gender, state or region indicators, equivalised household income, education categories, the number of failed comprehension-check attempts, and the randomised order of the additional games. Balance tests are reported in Appendix C; those tests show that sample balance is satisfactory in both countries. Standard errors are clustered at NUTS2 level.

	(1)	(2)	(3)	(4)
	Beliefs about:			
	Group Member's contribution		Administrator's honesty	
VARIABLES	Germany	Romania	Germany	Romania
Cond. Inefficiency	-0.036	-0.004	-4.361***	0.926
	(0.021)	(0.134)	(1.342)	(1.274)
Cond. Involuntary	-0.024	-0.259	-4.649***	-1.608
	(0.021)	(0.211)	(0.874)	(1.570)
Cond. Unconstrained	-0.050**	-0.323	-6.020***	-6.414***
	(0.019)	(0.171)	(0.907)	(1.659)
Constant	0.836***	5.233*	62.703***	58.017***
	(0.116)	(2.462)	(7.199)	(9.803)
Individual controls	✓	✓	✓	✓
Observations	3,210	1,998	3,210	1,998
R-squared	0.035	0.030	0.127	0.083

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Beliefs about group members' and administrators' behaviour in Germany and Romania in different experimental conditions

Table 2 shows that when interacting within an institution that leaves room for embezzlement of public funds (Condition Unconstrained), individuals indeed believe that administrators will make larger attempts at theft. The difference between the Unconstrained and Baseline conditions is large and statistically significant and this effect is true in both Romania and Germany, highlighting that our treatment worked as expected. In fact, in all treatments where there is a possibility of losses, participants in Germany are more distrustful of the administrator's behaviour compared to the treatment with no possibility of losses, even when theft or losses are forced by design. As for the beliefs about the contributions of others, the effects are different in Germany and Romania. In Germany, after exposure to a possibility of embezzlement, individuals believe other group members will contribute less than after exposure to no possibility of losses. In Romania, there is no such difference in the beliefs about other group members' behaviour between the conditions.

We continue by testing H1, which postulates that exposure to corruption generates less cooperation than exposure to institutional resource loss resulting from inefficiency. Because the hypothesis compares cooperation in the Unconstrained condition, which captures the possibility of embezzlement, with cooperation in the Inefficiency condition, the results are easiest to interpret when one is used as the reference category. We use the Unconstrained condition as the reference when comparing cooperation levels, measured by monetary contributions to the common pool. Tables 3 and 4 present the results for

Germany and Romania, respectively.

VARIABLES	(1) Contribution	(2) Contribution	(3) Contribution	(4) Contribution
Cond. Baseline	0.048	0.015	-0.008	-0.009
	(0.028)	(0.022)	(0.024)	(0.021)
Cond. Inefficiency	-0.066**	-0.075***	-0.081***	-0.081***
	(0.023)	(0.023)	(0.024)	(0.025)
Cond. Involuntary	0.032	0.015	0.019	0.011
	(0.026)	(0.021)	(0.023)	(0.021)
Belief GM coop.		0.657***		0.589***
		(0.016)		(0.014)
Belief Admin honesty			0.009***	0.005***
			(0.001)	(0.000)
Constant	0.841***	0.325***	0.314*	0.121
	(0.166)	(0.103)	(0.167)	(0.106)
Individual controls	✓	✓	✓	✓
Observations	3,210	3,210	3,210	3,210
R-squared	0.042	0.382	0.168	0.408

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Extent of contributions to the common pool in different experimental conditions in Germany

VARIABLES	(1) Contribution	(2) Contribution	(3) Contribution	(4) Contribution
Cond. Baseline	0.115	-0.122	-0.136	-0.179
	(0.199)	(0.149)	(0.193)	(0.147)
Cond. Inefficiency	0.024	-0.210	-0.263	-0.276*
	(0.160)	(0.149)	(0.164)	(0.143)
Cond. Involuntary	-0.103	-0.150	-0.291	-0.198
	(0.176)	(0.190)	(0.154)	(0.189)
Belief GM coop.		0.731***		0.705***
		(0.014)		(0.014)
Belief Admin honesty			0.039***	0.010***
			(0.003)	(0.003)
Constant	5.167	1.577	3.150	1.181
	(2.906)	(1.023)	(2.531)	(0.979)
Individual controls	✓	✓	✓	✓
Observations	1,998	1,998	1,998	1,998
R-squared	0.029	0.506	0.115	0.511

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Extent of contributions to the common pool in different experimental conditions in Romania

In the German sample, participants contributed significantly less under the Inefficiency condition than under the Unconstrained condition. This finding rejects H1 and instead indicates the opposite: institutional resource loss reduced cooperation more than exposure to the possibility of corruption. This result is robust when controlling for participants' beliefs about the behaviour of other Group Members and of the Administrator. The difference in contribution levels between the Unconstrained and Baseline conditions is statistically insignificant, suggesting that participants in the German sample responded more negatively to the resource-loss condition than to the embezzlement condition.

In Romania, the estimates point in the same direction as those obtained in Germany: participants contributed less under the Inefficiency condition than under the Unconstrained condition. The effect is statistically significant at the 10% level only after controlling for beliefs about the behaviour of other Group Members and the Administrator. Differences between the other conditions and the Unconstrained condition are statistically insignificant. The estimates from Romania are therefore smaller and considerably less precise than those from Germany. The difference in statistical precision may also partly reflect the smaller Romanian sample.

As a robustness check for the bounded nature of the contribution outcome, we re-estimate the four specifications using fractional-logit models after expressing contributions as shares of the endowment. The resulting average marginal effects closely track the OLS estimates and leave the main comparison between the Inefficiency and Unconstrained conditions unchanged. Appendix E reports the full comparison and the fractional-logit predicted treatment means.

Our additional analyses, presented in Appendix D, reveal no systematic treatment effects on trust and trustworthiness, personal or perceived social norms of cooperation, trust, and trustworthiness, or risk-taking across countries. One possible explanation for these results is respondent fatigue: by the time participants reached questions measuring these outcomes, they had completed a lengthy survey and worked through a substantial amount of video-based instructions, which may have reduced attention and attenuated treatment responsiveness. It could be, however, that these are not the mechanisms behind the effects of embezzlement and institutional resource loss on cooperation found in the main analysis and that there are other mechanisms in play. One such alternative could be, for instance, blame attribution, which we invite future research to investigate. One potential exception is belief about others' trustworthiness. We found that in the German sample, participants had lower perceptions of others' trustworthiness after exposure to the Unconstrained condition. This result is in line with those reported by Banerjee (2016) and highlights that some aspects of the trust mechanisms discussed in the theory section might be important for understanding cooperation in the presence of corruption. This finding, however, is not consistent across country contexts, and, therefore, needs further investigation.

6 Discussion

The findings qualify the expectation that deliberate violations of institutional integrity should be especially corrosive to cooperation. In the German sample, efficiency related failure to transform collective contributions into collective returns generated a stronger behavioural response than the possibility of private misappropriation. This suggests that citizens may respond not only to the moral character of an institutional failure, but also to whether institutions appear capable of safeguarding and effectively using collective resources.

One possible interpretation is that wrongdoing by an identifiable actor can be attributed to a particular source, whereas impersonal resource loss may be perceived as a broader institutional-performance failure whose cause is more difficult to locate or control. Alternatively, participants may react particularly negatively when collective resources are lost without producing either a public or an identifiable private beneficiary.

The estimated treatment effects differ between the two national samples: the negative effect associated with the Inefficiency condition is more pronounced in Germany and weaker in Romania. Because treatment assignment was randomized within each country, the corresponding within-country treatment effects have a causal interpretation, which is found to be heterogeneous across the two samples.

The pattern is consistent with the possibility that participants entered the experiment with different expectations shaped by their prior institutional experiences. Under this interpretation, the experimental conditions may have represented different departures from participants' ordinary expectations in the two countries, though other uncontrolled for country differences may be at play.

7 Conclusions

In this paper, we have developed theoretical arguments for and provided experimental tests of how people's cooperative behaviour may change under different institutional dysfunctionalities—institutional resource loss stemming from inefficiencies and a possibility of embezzlement of these resources. Because comparing institutional conditions across countries creates the problem of various confounders, we designed a novel behavioural game, aimed at capturing these institutional differences in experimental settings.

After performing two large online lab-like experiments in Germany and Romania, we found a clear trend. Our results show that institutional resource loss stemming from inefficiencies can reduce cooperation at least as strongly as a possibility of these resources being stolen. In Germany, participants cooperated less under institutional resource loss than when an Administrator can privately appropriate resources. The results in Romania show a corresponding pattern, but it is weaker and considerably less precisely estimated.

In our study, we examined one specific form of corruption, embezzlement of public funds, and one specific dimension of institutional inefficiency, impersonal resource loss, rather than the full range of corrupt practices and sources of inefficiency occuring in real-

world political systems. More research is needed on how other aspects of institutional quality affect people's willingness to contribute to public good, including other forms of corruption and inefficiency.

Our focus was on two settings with different levels institutional quality and the effects of the experimental conditions were not identical across the two contexts. This heterogeneity is consistent with accounts in which prior institutional experience shapes reactions to institutional dysfunctionalities, although cross-national comparisons do not allow for causal interpretations. Research implementing comparable experimental designs across a broader range of settings, or directly measuring the relevant prior experiences and expectations, is encouraged to further investigate other contextual mechanism systematically, and the robustness of these results.

We also invite future research to investigate further why institutional resource loss can generate such strong behavioural responses relative to corruption, including the psychological reactions triggered by the two institutional conditions. Potential mechanisms include differences in perceived responsibility, controllability, institutional competence, and fairness. These mechanisms can be tested using our the embezzlement games and its variants, which we invite scholars to use when studying the behavioural consequences of institutional quality in experimental settings.

References

Akbaş, M., Ariely, D., and Yuksel, S. (2019). When is inequality fair? An experiment on the effect of procedural justice and agency. *Journal of Economic Behavior & Organization*, 161:114–127.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, 126(4):556.

Attanasi, G., Rimbaud, C., and Villeval, M. C. (2019). Embezzlement and guilt aversion. *Journal of Economic Behavior & Organization*, 167:409–429.

Banerjee, R. (2016). Corruption, norm violation and decay in social capital. *Journal of Public Economics*, 137:14–27.

Banerjee, R., Boly, A., and Gillanders, R. (2022). Is corruption distasteful or just another cost of doing business? *Public Choice*, 190:33–51.

Barr, A. and Serra, D. (2010). Corruption and culture: An experimental analysis. *Journal of Public Economics*, 94(11):862–869.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10(1):122–142.

Bicchieri, C., Dimant, E., and Xiao, E. (2021). Deviant or wrong? The effects of norm information on the efficacy of punishment. *Journal of Economic Behavior & Organization*, 188:209–235.

Braccioli, F., Daniele, G., and Martinangeli, A. F. (2025). Breaking negative narratives: Long-term Social Progress and Trust in Institutions . *LEMMA Working Paper*.

Brennan, G., Eriksson, L., Goodin, R. E., and Southwood, N. (2013). *Explaining norms*. OUP UK.

Buso, I. M., Di Cagno, D., Ferrari, L., Larocca, V., Lorè, L., Marazzi, F., Panaccione, L., and Spadoni, L. (2021). Lab-like findings from online experiments. *Journal of the Economic Science Association*, 7(2):184–193.

Campos-Vazquez, R. M. and Mejia, L. A. (2016). Does corruption affect cooperation? a laboratory experiment. *Latin American Economic Review*, 25(1):5.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chen, X.-P., Au, W. T., and Komorita, S. (1996). Sequential choice in a step-level public goods dilemma: The effects of criticality and uncertainty. *Organizational Behavior and Human Decision Processes*, 65(1):37–47.

Clifford, S. and Jerit, J. (2014). Is there a cost to convenience? an experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, 1(2):120–131.

Crosetto, P. and Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Davidovic, D., Harring, N., and Jagers, S. C. (2020). The contingent effects of environmental concern and ideology: institutional context and people's willingness to pay environmental taxes. *Environmental Politics*, 29(4):674–696.

Dinesen, P. T. (2012). Parental transmission of trust or perceptions of institutional fairness: generalized trust of non-western immigrants in a high-trust society. *Comparative Politics*, 44(3):273–289.

Engl, F., Riedl, A., and Weber, R. (2021). Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *American Economic Journal: Microeconomics*, 13(4):261–299.

European Commission (2025). The vat gap. `https://taxation-customs.ec.europa.eu/taxation/vat/fight-against-vat-fraud/vat-gap_en`. Accessed: 16 January 2026.

Fisman, R. and Miguel, E. (2007). Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets | Journal of Political Economy: Vol 115, No 6. *Journal of Political Economy*, 115(6):1020–1049.

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

Grimes, M. (2006). Organizing consent: The role of procedural fairness in political trust and compliance. *European Journal of Political Research*, 45(2):285–315.

Harring, N. (2016). Reward or punish? understanding preferences toward economic or regulatory instruments in a cross-national perspective. *Political Studies*, 64(3):573–592.

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

Kampen, J. K., Van de Walle, S., and Bouckaert, G. (2006). Assessing the relation between satisfaction with public service delivery and trust in government: The impact of the predisposition of citizens toward government on evaluations of its performance. *Public Performance & Management Review*, 29(4):387–404.

Martinangeli, A. F., Povitkina, M., Jagers, S., and Rothstein, B. (2024). Institutional Quality Causes Generalized Trust: Experimental Evidence on Trusting under the Shadow of Doubt. *American Journal of Political Science*, 68(3). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12780.

Muthukrishna, M., Francois, P., Pourahmadi, S., and Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behaviour*, 1(7):0138.

Persson, A., Rothstein, B., and Teorell, J. (2013a). Why anticorruption reforms fail—systemic corruption as a collective action problem. *Governance*, 26(3):449–471.

Persson, A., Rothstein, B., and Teorell, J. (2013b). Why Anticorruption Reforms Fail—Systemic Corruption as a Collective Action Problem. *Governance*, 26(3):449–471. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0491.2012.01604.x.

Rajkumar, A. S. and Swaroop, V. (2008). Public spending and outcomes: Does governance matter? *Journal of development economics*, 86(1):96–111.

Rothstein, B. (2011). *The quality of government: Corruption, social trust, and inequality in international perspective*. University of Chicago Press.

Rothstein, B. (2013). Corruption and social trust: Why the fish rots from the head down. *Social Research*, 80(4):1009–1032.

Rothstein, B. and Eek, D. (2009). Political corruption and social trust: An experimental approach. *Rationality and society*, 21(1):81–112.

Rothstein, B. O. and Teorell, J. A. (2008). What is quality of government? a theory of impartial government institutions. *Governance*, 21(2):165–190.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer-Verlag, New York.

Spadaro, G., Molho, C., Van Prooijen, J.-W., Romano, A., Mosso, C. O., and Van Lange, P. A. M. (2023). Corrupt third parties undermine trust and prosocial behaviour between people. *Nature Human Behaviour*, 7(1):46–54.

Steinmo, S. and D'Attoma, J. (2022). *Willing to pay?: a reasonable choice approach*. Oxford University Press, Oxford.

Sun, Q., Molenmaker, W. E., Zhang, Y., Liu, Y., and van Dijk, E. (2022). Procedural fairness facilitates cooperative behavior by enhancing cooperative expectations. *Social and Personality Psychology Compass*, 16(12):e12711.

Tanzi, V. and Davoodi, H. (1998). Corruption, public investment, and growth. In *The welfare state, public investment, and growth: selected papers from the 53rd congress of the International Institute of Public Finance*, pages 41–60. Springer.

Tepe, M., Paetzel, F., Lorenz, J., and Lutz, M. (2021). Efficiency loss and support for income redistribution: evidence from a laboratory experiment. *Rationality and Society*, 33(3):313–340.

Transparency International (2025). What is corruption? Accessed: 2025-11-04.

Van de Walle, S. and Bouckaert, G. (2003). Public service performance and trust in government: The problem of causality. *International Journal of Public Administration*, 26(8–9):891–913.

Van Dijk, E., Wit, A., Wilke, H., and Budescu, D. V. (2004). *What We Know (and Do Not Know) about the Effects of Uncertainty on Behavior in Social Dilemmas*. Cambridge University Press.

Varieties of Democracy (V-Dem) Institute (2025). V-dem data analysis: Variable graph. Accessed: 2025-11-04.

Appendix

A Payoff implications of embezzlement and inefficiencies

Participants in groups of size n all choose how much to contribute to a collective good out of a personal endowment E. Each group member receives in return a fraction of their and others' contribution to the public good, but keeps the amount they have not contributed for themselves. In particular, the final payoff π_i of each individual $i \in G$, where G denotes the group, follows the rule

$$\pi_i = E - c_i + \alpha \sum_{j \in G} c_j = E - c_i + \alpha C, \tag{4}$$

with C denoting total group contributions, and $1/n < \alpha < 1$. The restriction on α models the tension between self- and collective interest faced by the individuals contributing to a public good. In particular, one Euro contributed by i increases i's payoff by an amount $\alpha < 1$. The other group members, however, also receive α as a result of i's contribution. The whole group thus benefits from each individual's contribution because the sum of all individuals' gains amounts to $n \cdot \alpha > n \cdot 1/n = 1$. Society (the group) therefore benefits most if everyone contributes the maximum possible amount to the common good, although each individual's self-interest is to keep his or her own endowment, let others contribute, and free ride on others' contributions.

Embezzlement and inefficiency are equivalent in terms of their payoff implications for non-Administrator Group Members: in both cases the amount available for redistribution is reduced. The aggregate implications differ, however, because resources privately appropriated by the Administrator remain part of group earnings, whereas resources removed by the software are lost. Embezzlement and institutional resource loss should therefore be distinguished conceptually even though they can impose the same direct payoff loss on non-Administrators. To see this, recall that αC represents an individual's return from the group's public good when total contributions are C. Denote by $R < C$ the amount removed before redistribution. Each non-Administrator's return is then $\alpha(C - R) < \alpha C$, implying an individual loss $L = \alpha R$. When R disappears through inefficiency, the total social-welfare loss is $SWL = n\alpha R$.

Compare this scenario with one in which instead of being inefficiently lost, the amount R is embezzled by the administrator. For the $n-1$ non-Administrators, the loss remains $L = \alpha R$. The administrator instead, while losing αR from redistribution, gains R from embezzlement, resulting in $L_a = \alpha R - R < 0$, a negative loss, i.e. a gain.[4] In this case, the total social welfare loss is computed as $\widehat{SWL} = (n-1)L + L_a$. Replacing the definitions of L and L_a, we obtain $\widehat{SWL} = (n-1)\alpha R + \alpha R - R = n\alpha R - R < SWL$. Losses in group earnings in the case of embezzlement can be considered, from a purely

[4]Recall that the administrator is part of the group and thus benefits from public-good provision. This assumption reflects reality: a tax administrator providing public goods, such as public roads, with public tax money also uses those roads themself. Similarly, a tax administrator embezzling part of the tax revenue will themself drive on poorer quality roads but will privately enjoy the stolen money.

accounting perspective, lower in size than in the case of inefficiencies. This result reflects the fact that, while not redistributed, R remains within the economy, though privately enjoyed by one individual instead of being redistributed. Crucially, while at the societal level the aggregate loss differs between inefficiencies and embezzlement, individual payoff consequences are identical and fixed at $L = \alpha R$ for non-administrators.

B Equilibrium in a modified game with embezzlement by an intermediary agent

We denote total group contributions as

$$C = \sum_{i \in G} c_i$$

The Administrator, k, is a member of G, and collects the group's contributions and redistributes them as in the standard public good game. We allow k to embezzle any fraction of C. The embezzled amount is R, with $0 \leq R \leq C$. If $R = 0$, all contributions are redistributed. If $R = Y$, the whole amount is embezzled. The payoffs of player $i \neq k$ are then given by

$$\pi_i = E - c_i + \alpha(Y - R).$$

Further, k's payoff is given by

$$\pi_k = E - c_k + \alpha(Y - R) + R.$$

Clearly, k maximises their payoff by setting $R = Y$. The Nash equilibrium remains that of zero contributions by player $i \neq k$ as in the standard public good game. The very same logic applies if R is lost due to inefficiencies instead of embezzled.

C Balance tests and sample composition

Tables C.1 and C.2 report covariate balance across the four experimental conditions in the German and Romanian estimation samples. The tables present condition-specific means and pairwise p-values comparing each non-baseline condition with the Baseline condition. The p-values are based on standard errors clustered at the experimental-group level. Table C.3 describes differences in sample composition between Germany and Romania. Because country was not randomly assigned, these comparisons are descriptive rather than randomization-balance tests.

Table C.1: Covariate balance across experimental conditions: Germany

	Treatment-arm means				Pairwise p-values		
Variable	Baseline	Inefficiency	Involuntary	Unconstrained	Ineff.–Base	Invol.–Base	Unconstr.–Base
Age 18–24	0.115	0.083	0.092	0.074	0.032	0.115	0.004
Age 25–34	0.145	0.177	0.156	0.152	0.075	0.542	0.670
Age 35–49	0.219	0.224	0.235	0.256	0.801	0.432	0.086
Age 50–64	0.292	0.282	0.266	0.280	0.656	0.216	0.565
Age 65 or older	0.229	0.234	0.251	0.238	0.785	0.261	0.648
Female	0.484	0.491	0.512	0.480	0.776	0.252	0.865
Other or diverse gender	0.002	0.000	0.000	0.003	0.157	0.157	0.980
Equivalized household income	1592.632	1571.869	1616.368	1620.700	0.602	0.559	0.484
Incorrect comprehension submissions	1.759	1.933	1.808	1.821	0.110	0.650	0.556
Order of additional tasks	2.462	2.525	2.474	2.456	0.454	0.885	0.943
Survey entry sequence	1538.299	1557.799	1544.520	1554.142	0.594	0.865	0.661
Education: no certificate	0.005	0.003	0.004	0.004	0.435	0.702	0.729
Education: lower or intermediate secondary	0.397	0.377	0.401	0.404	0.404	0.871	0.761
Education: upper secondary	0.276	0.260	0.268	0.272	0.450	0.709	0.845
Education: university degree	0.311	0.344	0.321	0.300	0.159	0.653	0.633
Education: doctorate	0.011	0.018	0.006	0.020	0.255	0.279	0.156
State: Bavaria	0.156	0.151	0.148	0.174	0.799	0.660	0.334
State: Berlin	0.034	0.047	0.053	0.044	0.199	0.072	0.309
State: Brandenburg	0.027	0.039	0.033	0.035	0.168	0.467	0.335
State: Bremen	0.010	0.018	0.004	0.009	0.170	0.128	0.832
State: Hamburg	0.020	0.018	0.032	0.018	0.784	0.115	0.762
State: Hesse	0.068	0.080	0.064	0.079	0.344	0.756	0.369
State: Mecklenburg–Western Pomerania	0.025	0.018	0.032	0.015	0.343	0.367	0.170
State: Lower Saxony	0.125	0.085	0.093	0.077	0.010	0.039	0.002
State: North Rhine–Westphalia	0.219	0.232	0.208	0.199	0.542	0.611	0.339
State: Rhineland–Palatinate	0.050	0.055	0.050	0.050	0.694	0.991	0.999
State: Saarland	0.015	0.006	0.012	0.013	0.098	0.662	0.709
State: Saxony	0.061	0.036	0.060	0.045	0.018	0.906	0.150
State: Saxony–Anhalt	0.018	0.027	0.023	0.037	0.273	0.501	0.028
State: Schleswig–Holstein	0.029	0.043	0.038	0.043	0.156	0.343	0.145
State: Thuringia	0.021	0.025	0.026	0.031	0.540	0.523	0.178
Observations	814	786	816	794	1,600	1,630	1,608

Notes: The sample consists of Group Members. The first four columns report condition-specific means. The final three columns report pairwise p-values comparing Inefficiency, Involuntary, and Unconstrained with Baseline. Standard errors are clustered at the experimental-group level.

Table C.2: Covariate balance across experimental conditions: Romania

	Treatment-arm means				Pairwise p-values		
Variable	Baseline	Inefficiency	Involuntary	Unconstrained	Ineff.–Base	Invol.–Base	Unconstr.–Base
Age 18–24	0.168	0.180	0.178	0.195	0.613	0.667	0.268
Age 25–34	0.202	0.190	0.188	0.203	0.649	0.600	0.962
Age 35–49	0.186	0.178	0.206	0.164	0.744	0.368	0.373
Age 50–64	0.188	0.198	0.174	0.172	0.685	0.590	0.533
Age 65 or older	0.257	0.254	0.253	0.266	0.912	0.871	0.750
Female	0.516	0.528	0.535	0.521	0.695	0.557	0.868
Other or diverse gender	0.002	0.000	0.000	0.000	0.318	0.318	0.318
Equivalized household income	1922.751	1954.876	1949.774	2003.263	0.542	0.618	0.151
Incorrect comprehension submissions	2.385	2.350	2.587	2.483	0.819	0.206	0.548
Order of additional tasks	2.595	2.544	2.597	2.588	0.638	0.982	0.953
Survey entry sequence	1643.136	1632.194	1633.543	1655.594	0.853	0.872	0.834
Education: no certificate	0.000	0.002	0.004	0.000	0.318	0.157	
Education: primary	0.006	0.008	0.012	0.004	0.693	0.305	0.674
Education: lower secondary	0.032	0.028	0.026	0.030	0.732	0.593	0.912
Education: upper secondary, level 3	0.115	0.116	0.104	0.079	0.947	0.593	0.061
Education: upper secondary, baccalaureate	0.289	0.298	0.257	0.298	0.746	0.272	0.741
Education: short university or bachelor's	0.336	0.336	0.355	0.365	0.999	0.542	0.340
Education: master's	0.204	0.194	0.222	0.211	0.697	0.446	0.767
Education: doctorate	0.020	0.018	0.020	0.012	0.836	0.975	0.334
Region: South-East	0.138	0.124	0.148	0.132	0.506	0.657	0.766
Region: South-Muntenia	0.144	0.156	0.162	0.136	0.594	0.432	0.709
Region: South-West Oltenia	0.099	0.090	0.096	0.089	0.631	0.884	0.589
Region: West	0.087	0.110	0.080	0.079	0.238	0.707	0.665
Region: North-West	0.115	0.126	0.098	0.116	0.572	0.386	0.961
Region: Centre	0.097	0.110	0.134	0.114	0.483	0.057	0.398
Region: Bucharest–Ilfov	0.144	0.128	0.146	0.160	0.444	0.925	0.478
Observations	506	500	499	493	1,006	1,005	999

Notes: The sample consists of Group Members. The first four columns report condition-specific means. The final three columns report pairwise p-values comparing Inefficiency, Involuntary, and Unconstrained with Baseline. Standard errors are clustered at the experimental-group level.

Table C.3: Sample composition in Germany and Romania

Variable	Germany	Romania	Difference / p-value	Std. difference
Age 18–24	0.091	0.180	0.089	0.185
	(0.288)	(0.384)	0.000	
Age 25–34	0.157	0.196	0.038	0.071
	(0.364)	(0.397)	0.001	
Age 35–49	0.233	0.184	-0.050	-0.087
	(0.423)	(0.387)	0.000	
Age 50–64	0.280	0.183	-0.097	-0.163
	(0.449)	(0.387)	0.000	
Age 65 or older	0.238	0.257	0.019	0.032
	(0.426)	(0.437)	0.110	
Female	0.492	0.525	0.033	0.047
	(0.500)	(0.499)	0.021	
Other or diverse gender	0.001	0.001	-0.001	-0.018
	(0.035)	(0.022)	0.351	
University degree	0.332	0.573	0.241	0.352
	(0.471)	(0.495)	0.000	
Incorrect comprehension submissions	1.829	2.451	0.622	0.190
	(2.155)	(2.470)	0.000	
Order of additional tasks	2.479	2.581	0.102	0.043
	(1.698)	(1.690)	0.038	
Observations	3,210	1,998	5,208	

Notes: The table reports country-specific means, with standard deviations in parentheses. The difference is Romania minus Germany; its p-value appears on the following line. The final column reports the standardized difference. These country comparisons are descriptive because country was not randomly assigned.

D Post-treatment trust, norms, and risk outcomes

D.1 Elicitation of the additional outcomes

The trust, norm, and risk modules were administered after the public-good game and before the final demographic questions. Their order was randomised, and each module was explained only when it began.

Trust Game. We elicited trust and trustworthiness with the standard Trust Game (Berg et al., 1995). Each participant made decisions in both roles, in random order. As Sender, participants chose how much of their country-specific endowment (2 euros in Germany and 10 lei in Romania) to send in one of eleven equally spaced amounts; the Receiver obtained twice the amount sent. As Receiver, participants used the strategy method to state, for every possible transfer, how much to return from their own endowment plus the doubled transfer. One randomly selected matching direction determined payment. Trust is measured by the share sent and trustworthiness by the mean share returned across the eleven contingent decisions.

Norm elicitation. Personal norms were measured by asking respondents to rate, on a four-point scale from "strongly disagree" to "strongly agree," the statements "Everyone should cooperate with others in the general interest," "Everyone should trust others," and "No one should betray the trust that others place in them." Perceived social norms were elicited by asking respondents to predict the modal response of the other participants to each statement; each correct prediction earned 60 cents.

Bomb Risk Elicitation Task. Risk taking was measured with the dynamic Bomb Risk Elicitation Task (BRET; Crosetto and Filippin, 2013). One bomb was hidden among 64 boxes, which were collected in random order at one box per second until the participant stopped the task. Each box collected increased the potential payoff by a fixed amount, but collecting the bomb reduced the task payoff to zero. The number of boxes collected therefore provides an increasing measure of risk-taking.

We examine whether the experimental conditions affected behaviour in the additional games or respondents' personal and perceived social norms. To make the outcomes comparable across Germany and Romania, all dependent variables are expressed on a common zero-to-one scale. Trust is the share of the country-specific sender endowment sent; trustworthiness is the mean share returned out of the receiver's total available funds across the eleven contingent decisions; the norm measures are divided by their zero-to-three range; and risk-taking is the share of the 64 boxes collected. We estimate OLS models separately by country, with the Unconstrained condition as the reference category and standard errors clustered at the NUTS2 level.

All specifications include the same controls as the main analyses: age categories, gender, NUTS2 indicators, equivalised household income, education categories, the number of failed comprehension-check attempts, and the randomized order of the additional games. Column (1) includes treatment indicators and these controls; columns (2) and (3) additionally control for beliefs about Group Members' contributions and the Administrator's honesty, respectively; column (4) includes both belief measures.

D.2 Trust game

Table D.1 reports treatment effects on trust and trustworthiness. In Germany, the Baseline and Involuntary estimates are positive in the specifications without both belief measures, whereas in the Romanian sample, the estimates are smaller and imprecise. The overall pattern is sensitive to the inclusion of the belief controls and does not indicate a uniform treatment effect across countries.

Table D.1: Treatment effects on trust-game outcomes

	(1)	(2)	(3)	(4)
Germany: Trust sent				
Cond. Baseline	0.026**	0.016	0.008	0.008
	(0.011)	(0.012)	(0.012)	(0.012)
Cond. Inefficiency	0.020	0.017	0.015	0.015
	(0.012)	(0.013)	(0.012)	(0.013)
Cond. Involuntary	0.040**	0.035*	0.036**	0.033*
	(0.015)	(0.017)	(0.016)	(0.018)
Observations	3,210	3,210	3,210	3,210
R-squared	0.030	0.147	0.073	0.155
Germany: Trustworthiness returned				
Cond. Baseline	0.016**	0.013**	0.011*	0.011*
	(0.006)	(0.006)	(0.005)	(0.006)
Cond. Inefficiency	0.013	0.012	0.012	0.012
	(0.010)	(0.011)	(0.011)	(0.011)
Cond. Involuntary	0.023**	0.021*	0.021**	0.020*
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	3,210	3,210	3,210	3,210
R-squared	0.029	0.063	0.038	0.064
Romania: Trust sent				
Cond. Baseline	0.021	0.008	0.005	0.003
	(0.023)	(0.020)	(0.025)	(0.021)
Cond. Inefficiency	0.003	-0.010	-0.015	-0.015
	(0.028)	(0.027)	(0.029)	(0.028)
Cond. Involuntary	-0.004	-0.007	-0.016	-0.010
	(0.023)	(0.027)	(0.023)	(0.027)
Observations	1,998	1,998	1,998	1,998
R-squared	0.025	0.197	0.062	0.201
Romania: Trustworthiness returned				
Cond. Baseline	0.015	0.009	0.009	0.008
	(0.008)	(0.007)	(0.008)	(0.007)
Cond. Inefficiency	0.008	0.003	0.001	0.001
	(0.012)	(0.012)	(0.013)	(0.013)
Cond. Involuntary	-0.018	-0.019	-0.023	-0.021
	(0.017)	(0.017)	(0.016)	(0.017)
Observations	1,998	1,998	1,998	1,998
R-squared	0.023	0.075	0.034	0.076
Individual controls	✓	✓	✓	✓
Belief GM coop.		✓		✓
Belief Admin honesty			✓	✓

Notes: OLS coefficients relative to the Unconstrained condition. Outcomes range from zero to one. Standard errors clustered at the NUTS2 level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (uncorrected).

D.3 Norms

Tables D.2 and D.3 report effects on personal and perceived social norms of cooperation, trust, and trustworthiness. Most coefficients are small and do not form a consistent pattern across conditions and countries. The clearest exception is a positive association between being exposed to the Baseline and Involuntary conditions and perceived trustworthiness norms in Germany; the treatment coefficients for the other norm outcomes

are generally less stable across specifications.

Table D.2: Treatment effects on personal norms

	(1)	(2)	(3)	(4)
Germany: Cooperation				
Cond. Baseline	0.008	0.005	0.001	0.001
	(0.007)	(0.007)	(0.008)	(0.008)
Cond. Inefficiency	-0.005	-0.006	-0.007	-0.007
	(0.009)	(0.009)	(0.009)	(0.010)
Cond. Involuntary	-0.007	-0.008	-0.008	-0.009
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	3,210	3,210	3,210	3,210
R-squared	0.029	0.053	0.045	0.059
Germany: Trust				
Cond. Baseline	0.011	0.008	0.003	0.003
	(0.012)	(0.011)	(0.011)	(0.011)
Cond. Inefficiency	0.011	0.010	0.008	0.008
	(0.014)	(0.014)	(0.014)	(0.014)
Cond. Involuntary	0.019	0.017	0.017	0.016
	(0.014)	(0.013)	(0.014)	(0.013)
Observations	3,210	3,210	3,210	3,210
R-squared	0.038	0.056	0.051	0.061
Germany: Trustworthiness				
Cond. Baseline	0.017**	0.015**	0.010	0.010
	(0.007)	(0.006)	(0.006)	(0.006)
Cond. Inefficiency	-0.016	-0.016	-0.018*	-0.018*
	(0.010)	(0.010)	(0.010)	(0.010)
Cond. Involuntary	0.003	0.002	0.002	0.001
	(0.011)	(0.011)	(0.010)	(0.011)
Observations	3,210	3,210	3,210	3,210
R-squared	0.024	0.033	0.043	0.045
Romania: Cooperation				
Cond. Baseline	0.002	-0.001	-0.006	-0.007
	(0.019)	(0.017)	(0.019)	(0.018)
Cond. Inefficiency	-0.013	-0.016	-0.023*	-0.023*
	(0.013)	(0.012)	(0.012)	(0.012)
Cond. Involuntary	0.002	0.002	-0.004	-0.003
	(0.011)	(0.011)	(0.011)	(0.010)
Observations	1,998	1,998	1,998	1,998
R-squared	0.028	0.053	0.058	0.070
Romania: Trust				
Cond. Baseline	-0.014	-0.017	-0.019	-0.020
	(0.013)	(0.011)	(0.012)	(0.011)
Cond. Inefficiency	-0.008	-0.012	-0.015	-0.015
	(0.010)	(0.010)	(0.010)	(0.009)
Cond. Involuntary	-0.012	-0.013	-0.017	-0.015
	(0.012)	(0.014)	(0.012)	(0.013)
Observations	1,998	1,998	1,998	1,998
R-squared	0.038	0.060	0.046	0.062
Romania: Trustworthiness				
Cond. Baseline	-0.016	-0.018	-0.022*	-0.022*
	(0.012)	(0.012)	(0.011)	(0.011)
Cond. Inefficiency	-0.015	-0.017	-0.022*	-0.022*
	(0.011)	(0.010)	(0.011)	(0.011)
Cond. Involuntary	-0.014	-0.014*	-0.019*	-0.018*
	(0.008)	(0.008)	(0.008)	(0.008)
Observations	1,998	1,998	1,998	1,998
R-squared	0.023	0.033	0.038	0.041
Individual controls	✓	✓	✓	✓
Belief GM coop.		✓		✓
Belief Admin honesty			✓	✓

Notes: OLS coefficients relative to the Unconstrained condition. Outcomes range from zero to one. Standard errors clustered at the NUTS2 level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (uncorrected).

Table D.3: Treatment effects on perceived social norms

	(1)	(2)	(3)	(4)
Germany: Cooperation				
Cond. Baseline	0.023*	0.020	0.018	0.018
	(0.011)	(0.011)	(0.012)	(0.012)
Cond. Inefficiency	0.012	0.012	0.011	0.011
	(0.008)	(0.008)	(0.008)	(0.008)
Cond. Involuntary	0.019**	0.018*	0.018*	0.017*
	(0.008)	(0.009)	(0.009)	(0.009)
Observations	3,210	3,210	3,210	3,210
R-squared	0.030	0.046	0.036	0.047
Germany: Trust				
Cond. Baseline	0.009	0.007	0.004	0.004
	(0.010)	(0.010)	(0.011)	(0.010)
Cond. Inefficiency	-0.008	-0.008	-0.009	-0.009
	(0.011)	(0.010)	(0.011)	(0.011)
Cond. Involuntary	0.010	0.009	0.009	0.008
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	3,210	3,210	3,210	3,210
R-squared	0.048	0.057	0.052	0.059
Germany: Trustworthiness				
Cond. Baseline	0.036***	0.034***	0.032***	0.032***
	(0.011)	(0.010)	(0.010)	(0.010)
Cond. Inefficiency	0.007	0.006	0.006	0.006
	(0.010)	(0.009)	(0.010)	(0.009)
Cond. Involuntary	0.027**	0.026**	0.026**	0.025**
	(0.011)	(0.011)	(0.011)	(0.011)
Observations	3,210	3,210	3,210	3,210
R-squared	0.022	0.031	0.027	0.033
Romania: Cooperation				
Cond. Baseline	0.005	0.002	0.000	-0.000
	(0.019)	(0.018)	(0.019)	(0.018)
Cond. Inefficiency	0.005	0.002	-0.002	-0.002
	(0.009)	(0.009)	(0.010)	(0.010)
Cond. Involuntary	-0.001	-0.001	-0.005	-0.003
	(0.008)	(0.006)	(0.008)	(0.006)
Observations	1,998	1,998	1,998	1,998
R-squared	0.037	0.057	0.047	0.060
Romania: Trust				
Cond. Baseline	0.015	0.011	0.009	0.009
	(0.013)	(0.013)	(0.014)	(0.014)
Cond. Inefficiency	0.015	0.011	0.008	0.008
	(0.016)	(0.017)	(0.017)	(0.017)
Cond. Involuntary	0.021	0.020	0.017	0.018
	(0.015)	(0.014)	(0.014)	(0.014)
Observations	1,998	1,998	1,998	1,998
R-squared	0.043	0.062	0.051	0.064
Romania: Trustworthiness				
Cond. Baseline	0.015	0.013	0.009	0.009
	(0.010)	(0.010)	(0.010)	(0.009)
Cond. Inefficiency	0.016	0.014	0.009	0.009
	(0.016)	(0.016)	(0.017)	(0.017)
Cond. Involuntary	0.011	0.011	0.007	0.007
	(0.011)	(0.010)	(0.011)	(0.011)
Observations	1,998	1,998	1,998	1,998
R-squared	0.040	0.049	0.051	0.055
Individual controls	✓	✓	✓	✓
Belief GM coop.		✓		✓
Belief Admin honesty			✓	✓

Notes: OLS coefficients relative to the Unconstrained condition. Outcomes range from zero to one. Standard errors clustered at the NUTS2 level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (uncorrected).

D.4 Risk elicitation

Table D.4 shows no systematic treatment effects on risk taking in either country, and this conclusion is unchanged when either or both belief measures are included.

Table D.4: Treatment effects on risk taking

	(1)	(2)	(3)	(4)
Germany: Risk taking				
Cond. Baseline	0.005	0.004	0.005	0.005
	(0.009)	(0.009)	(0.009)	(0.009)
Cond. Inefficiency	-0.004	-0.005	-0.004	-0.004
	(0.008)	(0.008)	(0.008)	(0.008)
Cond. Involuntary	-0.009	-0.009	-0.009	-0.009
	(0.010)	(0.010)	(0.010)	(0.010)
Observations	3,210	3,210	3,210	3,210
R-squared	0.025	0.027	0.025	0.027
Romania: Risk taking				
Cond. Baseline	-0.002	-0.004	0.002	0.002
	(0.021)	(0.022)	(0.022)	(0.021)
Cond. Inefficiency	0.005	0.004	0.011	0.010
	(0.022)	(0.022)	(0.023)	(0.022)
Cond. Involuntary	0.023	0.023	0.026	0.027
	(0.025)	(0.025)	(0.025)	(0.025)
Observations	1,998	1,998	1,998	1,998
R-squared	0.028	0.031	0.032	0.038
Individual controls	✓	✓	✓	✓
Belief GM coop.		✓		✓
Belief Admin honesty			✓	✓

Notes: OLS coefficients relative to the Unconstrained condition. The outcome is the share of the 64 boxes collected. Standard errors clustered at the NUTS2 level are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (uncorrected).

E Robustness checks

Our main specifications use OLS, whose treatment coefficients directly estimate differences in mean contributions under random assignment. Because the contribution outcome is bounded, we assess robustness using fractional-logit models. We express contributions as shares of the endowment, so that the dependent variable ranges from zero to one in both samples. We then re-estimate the four country-specific specifications used in the main analysis. All specifications include age categories, gender, state or region indicators, equivalised household income, education categories, the number of failed comprehension-check attempts, and the randomised order of the additional games. Standard errors are clustered at the NUTS2 level. Because coefficients from a nonlinear model are not directly comparable with OLS coefficients, we report average marginal effects.

Table E.1 compares the OLS treatment coefficients, estimated on the rescaled outcome, with the fractional-logit average marginal effects. In Germany, the estimated difference between the Inefficiency and Unconstrained conditions ranges from approxi-

mately −0.033 to −0.038 in the fractional-logit models and is statistically significant at the one-percent level in every specification. These estimates closely match the corresponding OLS estimates. In Romania, the difference remains smaller and less precisely estimated. It is statistically significant at the ten-percent level only in the fourth specification, which controls for both beliefs about other Group Members' contributions and beliefs about the Administrator's behaviour (average marginal effect = −0.025, $p = 0.072$). Thus, the fractional-logit results preserve the central pattern reported in the main analysis. Two ancillary contrasts become marginally significant at the ten-percent level under fractional logit—Baseline versus Unconstrained in the first German specification and Involuntary versus Unconstrained in the third Romanian specification—but neither affects the central comparison between Inefficiency and Unconstrained.

Table E.1: OLS and fractional-logit estimates of treatment differences

	Specification 1		Specification 2		Specification 3		Specification 4	
Comparison	OLS	Fractional	OLS	Fractional	OLS	Fractional	OLS	Fractional
Germany								
Baseline – Unconstrained	0.024	0.024*	0.008	0.008	-0.004	-0.001	-0.004	-0.002
	(0.014)	(0.014)	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)
Inefficiency – Unconstrained	-0.033**	-0.033***	-0.037***	-0.035***	-0.041***	-0.038***	-0.041***	-0.038***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Involuntary – Unconstrained	0.016	0.016	0.008	0.009	0.010	0.012	0.005	0.007
	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Romania								
Baseline – Unconstrained	0.011	0.012	-0.012	-0.011	-0.014	-0.013	-0.018	-0.016
	(0.020)	(0.020)	(0.015)	(0.014)	(0.019)	(0.019)	(0.015)	(0.014)
Inefficiency – Unconstrained	0.002	0.003	-0.021	-0.019	-0.026	-0.025	-0.028*	-0.025*
	(0.016)	(0.016)	(0.015)	(0.014)	(0.016)	(0.016)	(0.014)	(0.014)
Involuntary – Unconstrained	-0.010	-0.010	-0.015	-0.017	-0.029	-0.028*	-0.020	-0.021
	(0.018)	(0.017)	(0.019)	(0.018)	(0.015)	(0.015)	(0.019)	(0.019)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GM belief	No	No	Yes	Yes	No	No	Yes	Yes
Administrator belief	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the contribution expressed as a share of the endowment. OLS entries are treatment coefficients and fractional-logit entries are average marginal effects; all comparisons are relative to the Unconstrained condition. All specifications include the individual covariates used in the main analysis. Specification 2 adds beliefs about Group Members' contributions, Specification 3 adds beliefs about the Administrator's behaviour, and Specification 4 includes both belief measures. Standard errors, clustered at the NUTS2 level, are in parentheses. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table E.2 reports fractional-logit predicted mean contribution shares for each experimental condition. In Germany, the predicted contribution share in the Inefficiency condition is consistently lower than in the Unconstrained condition across all four specifications. In Romania, the corresponding difference is smaller and becomes apparent only after the belief measures are included, consistent with the weaker Romanian results in the main analysis.

Table E.2: Fractional-logit predicted contribution shares by experimental condition

Condition	Specification 1	Specification 2	Specification 3	Specification 4
Germany				
Baseline	0.718	0.708	0.701	0.702
	(0.009)	(0.006)	(0.006)	(0.006)
Inefficiency	0.661	0.665	0.665	0.666
	(0.006)	(0.007)	(0.007)	(0.007)
Involuntary	0.710	0.709	0.714	0.711
	(0.007)	(0.006)	(0.007)	(0.006)
Unconstrained	0.694	0.701	0.703	0.704
	(0.008)	(0.007)	(0.008)	(0.007)
Romania				
Baseline	0.611	0.601	0.604	0.600
	(0.009)	(0.007)	(0.009)	(0.007)
Inefficiency	0.602	0.593	0.592	0.591
	(0.007)	(0.006)	(0.007)	(0.006)
Involuntary	0.589	0.595	0.589	0.595
	(0.009)	(0.009)	(0.009)	(0.010)
Unconstrained	0.599	0.612	0.617	0.616
	(0.012)	(0.011)	(0.011)	(0.011)

Notes: Entries are average predicted contribution shares from the fractional-logit models; standard errors, clustered at the NUTS2 level, are in parentheses. All specifications include the individual covariates used in the main analysis. Specification 2 adds beliefs about Group Members' contributions, Specification 3 adds beliefs about the Administrator's behaviour, and Specification 4 includes both belief measures.

F The survey

The study was programmed on oTree (Chen et al., 2016) and fielded in the local languages (German and Romanian). Below is a full transcription of the survey the participants went through, translated into English. The instructions were delivered to the participants in two ways. First, YouTube videos, created with animations in PowerPoint, were embedded within the survey environment to explain the structure of the survey, the games, the payment schemes and the matching protocol, and screen-out procedures for low-quality answers. Clickable hyperlinks to the English version of the videos are embedded within the survey transcription.

All videos could be replayed by the participants at specific checkpoints within the survey (these are marked within the survey transcription). Second, at the checkpoints, the participants could click through stills of the videos in the PNG format, at their own pace.

Notes added by the authors for readers of the appendix, which were not shown to the participants, appear as indented blue paragraphs headed "Note for the reader." Each survey question is numbered separately within its substantive block, and numbered response categories appear immediately below it. Interface controls such as "Continue,"

"Solve," and "Finish" are presented separately as bulleted lists.

F.1 Introduction

Informed consent was included with the introduction video, detailing the anonymity of the data, the possibility for the participant to withdraw consent at any moment by quitting the survey, and that the participant could earn an additional compensation as part of the behavioural incentivisation scheme (in addition to the remuneration already offered by the survey company) by completing the survey. The participants were also warned that no compensation would accrue to those who would not complete the survey. The survey consisted of four incentivised games, only one of which was randomly selected to be paid to the participants, who were informed of this payment scheme from the outset.

Introduction video: `https://youtu.be/GFW0AWRkHK8`

- No, I want to leave the survey
- Continue

Please answer the following questions.

1What is your age?

18–29

30–39

40–49

50–59

60–69

70+

2What is your gender?

Male

Female

Other

3What is your state or region of residence?

Country-specific list of German states or Romanian regions

4What is the size of your household? Free numerical field

Should the survey hit the quota limits for representativeness (geography, age, gender), the participant would be screened out with this page.

Thank you for participating. The session is already full.

Please wait while we redirect you to the panel website.

F.2 Public good game explanation

Public goods game explanation: https://youtu.be/09fZe-qZC-M

- Continue

The administrators' role was explained on this page.
Administrator explanationhttps://youtu.be/STx6QTOLYP0

- Continue

The participants were randomised into one of the experimental conditions at this stage. Each participant was randomised into one condition only, and hence only watched one of the following videos.
Experimental condition – baseline: https://youtu.be/py0-u6UNKlc
Experimental condition – involuntary: https://youtu.be/hpLkQjiQxs8
Experimental condition – inefficiency: https://youtu.be/vZNZUtf0AGI
Experimental condition – unconstrained: https://youtu.be/-QndeO5Km0o

- Continue

This is a checkpoint page, where the participants could re-watch the videos they had been shown (and only those), or click through the video stills

On this page you can re-watch the videos you saw earlier, or click on "Continue" to proceed.

- Continue

- Re-watchable video 1
- Re-watchable video 2
- Re-watchable condition-specific video
- Video stills

Control questions
On this page, we included control questions to gauge the respondents' understanding of the public good game payoff computations. Because our interest was in ensuring understanding rather than capturing the respondents' degree of understanding, we included a clickable button to display solutions. We nevertheless included a counter of the number of times a participant would submit the page with incorrect answers, and

capped the number of attempts at 10. After 10 incorrect attempts at submitting the form, the participant was screened out (this occurred only once).

Before we proceed:

It is **important** that the calculation of earnings is clear to you.
Please answer the following questions to check your understanding. Use your calculator if needed.
Click on "Solve" to display the solutions.

Here is a reminder of the calculation:

Your earnings = your 2 euros − your contribution + the **double of the group's total contribution**, **equally distributed** among all 3 group members

Your earnings = your 2 euros − your contribution + ((your contribution + the contribution of the others) × 2) / 3

Be careful! The arrow keys on your keyboard can change your entries!

1Imagine that everyone, including you, contributes **nothing** to the project. How much does **each individual** group member earn?

Higher than that of the other group members and equal to 2€

Higher than that of the other group members and equal to 4€

The same as that of the other group members and equal to 2€

Lower than that of the other group members and equal to 2€

Lower than that of the other group members and equal to 0€

The same as that of the other group members and equal to 4€

- Solve Question 1

  Only displayed if "Solve" is clicked Solution: Everyone keeps their own **2 euros**.

2Imagine that everyone, including you, contributes **their entire endowment** to the project. How much does **each individual** group member earn?

Higher than that of the other group members and equal to 2€

Higher than that of the other group members and equal to 4€

The same as that of the other group members and equal to 4€

Lower than that of the other group members and equal to 2€

Lower than that of the other group members and equal to 0€

The same as that of the other group members and equal to 2€

- Solve Question 2

Only displayed if "Solve" is clicked Solution: Everyone contributes 2€. There are 6€ in the project, which doubled become 12 euros. They are divided equally. Therefore, each person in the group receives **2 − 2 + ((2 + 2 + 2) × 2) / 3 = 12/3 = 4 euros**.

3aImagine that the **other two group members** together contribute 2€ to the project. How much will **you** earn if **you add 0€** to the 2€?

Higher than that of the other group members

Lower than that of the other group members

The same as that of the other group members

- Solve Question 3a

  Only displayed if "Solve" is clicked Solution: There are 2€ in the project, which doubled become 4€. You receive 1/3 of that in addition to the 2€ you kept, so you earn: **2 − 0 + (2 × 2) / 3 = 2 + 4 / 3 = 2 + 1.3 = 3.3€**.

3bHow much will **you** earn if **you add 2€** to the 2€?

Higher than that of the other group members

Lower than that of the other group members

The same as that of the other group members

- Solve Question 3b

  Only displayed if "Solve" is clicked Solution: There are 4€ in the project, which doubled become 8€. You receive 1/3 of that in addition to the 0€ you kept, so you earn **2 − 2 + ((2 + 2) × 2) / 3 = 0 + (4 × 2) / 3 = 0 + 8 / 3 = 0 + 2.6€**.

4aImagine that **you** contribute 1€ to the project. How much will **you** earn if **the others together add 0€** to the 1€?

Higher than that of the other group members

Lower than that of the other group members

The same as that of the other group members

- Solve Question 4a

  Only displayed if "Solve" is clicked Solution: There is 1€ in the project, which doubled becomes 2€. You receive 1/3 of that in addition to the 1€ you kept, so you earn **1 + 2 / 3 = 1 + 0.6 = 1.6 euros**.

4bHow much will **you** earn if **the others together add 2€** to the 1€?

Higher than that of the other group members

Lower than that of the other group members

The same as that of the other group members

- Solve Question 4b

  Only displayed if "Solve" is clicked Solution: There are 3€ in the project, which doubled become 6€. You receive 1/3 of that in addition to the 1€ you kept, so you earn $\mathbf{2 - 1 + ((1 + 2) \times 2) / 3 = 1 + (3 \times 2) / 3 = 1 + 6/3 = 1 + 2 = 3€}$.

- Continue

Contribute

The participants could submit their contribution to the public good from 0 to 2 euros in steps of 1 euro cent. They could do so by moving an initially invisible cursor on a slider bar. Real-time computations would display both the amount contributed by the participant and what they would keep for themself.

5How much will you contribute to the project (from 0 to 200)?

- Continue

Redistribute

This page is only displayed to the administrators.

6You were selected to collect the contributions. How much of them will you declare (from 0% to 100%)?

Click on the blue bars to display the sliders, and move them to select your answer.

- Continue

Redistribute (correction page)

Conditions Baseline, Involuntary and Inefficiency require the Administrator to redistribute the full amount contributed by the group. Should they attempt to embezzle any fraction of the contributions, they would land on this page, where they were asked to correct their answer.

You did not redistribute the full amount collected by the group. Please revise your entry.

7You were selected to collect the contributions. How much of them will you declare (from 0% to 100%)?

Click on the blue bars to display the sliders, and move them to select your answer.

- Continue

Belief elicitation for non-administrator, contribution question shown first

The next two pages display the belief elicitation questions for non-administrator group members – about the contribution behaviour of the other group member, and about the administrator's embezzlement behaviour – displayed in random order.

Please answer the following questions.
We pay you an **additional amount** if you guess correctly.
The closer your estimate is to the true value, the more you receive, up to 0.5€ per estimate.

8How much do you think the other group member (not the administrator) contributed?

Click on the blue bars to display the sliders, and move them to select your answer.

- Continue

Please answer the following questions.
We pay you an **additional amount** if you guess correctly.
The closer your estimate is to the true value, the more you receive, up to 0.5€ per estimate.

9How much of the group contributions do you think the administrator declared for redistribution (from 0% to 100%)?

Click on the blue bars to display the sliders, and move them to select your answer.

- Continue

Belief elicitation for administrator

The next questions are only displayed to Administrators to elicit their beliefs about the group's contribution behaviour.

Please answer the following questions.
We pay you an **additional amount** if you guess correctly.
The closer your estimate is to the true value, the more you receive, up to 5 points per estimate.

10How much do you think the other group members contributed on average (rounded to the nearest 0.1 euro)?

Click on the blue bars to display the sliders, and move them to select your answer.

- Continue

F.3 Additional games

Three further sets of behavioural games followed in random order: trust game, risk elicitation, and an incentivised opinion matching norm elicitation strategy. These were intended to more deeply investigate the mechanisms uncovered with the public good game, which was always played first. These games were not disclosed to the participants before the end of the public good game section of the study. Because the data collected from these games does not return any informative insight we postpone their description and transcription to Appendix G, and omit their discussion in the interest of space.

F.4 Further demographics and corruption questions

1How many people live in your household (including yourself)? Free numerical field
2How many children (under 18 years old) live in your household? Free numerical field
3What was your household's **monthly net income** last year? By this we mean the total amount resulting from wages, salaries, income from self-employment, retirement pensions, public benefits, income from rent and leasing, housing benefits, child benefits, and other income, **after deduction of taxes and social security contributions**.

<500€

500€–1500€

1501€–2000€

2001€–2500€

2501€–4499€

4500€ +

4What is your highest educational qualification (general or vocational)? Response labels were adapted to each country's education system; the harmonised English categories are shown here.

No school-leaving qualification

Lower secondary, intermediate secondary, or comparable qualification

High school/professional education or comparable qualification

Completed university or university of applied sciences degree (Bachelor, Master, etc.)

PhD

5What is your current employment status?

Employed full-time

Employed part-time

Self-employed or business owner

I am unemployed and looking for work

I am a pupil / student / trainee / apprentice / retraining participant

Retired

I am currently not employed and not looking for work

6How often do you take part in surveys?

Never

Sometimes

Often

Very often

7How many surveys, including this one, have you completed today? Free numerical field

- Continue

Corruption perceptions and experience The third corruption question was conditionally required: it had to be left blank if the participant answered "No" to the second corruption question, and had to be answered if the participant answered "Yes".

1How often do you believe employees in the public sector in your region embezzle public funds or other state resources for personal or family purposes?

OFTEN. Public sector employees are responsible custodians of selected public funds, but treat the rest like personal property.

ABOUT HALF THE TIME. The probability that public sector employees handle selected public funds responsibly is about equal to the probability that they treat them like personal property.

OCCASIONALLY. Public sector employees handle most public resources responsibly, but treat selected others like personal property.

NEVER OR ALMOST NEVER. Public sector employees almost always handle public funds responsibly and keep them separate from personal or family property.

2Have you ever had direct experience with corruption in your region?

Yes

No

3How often would you say you have had direct experiences with corruption in your region in the last 5 years?

Sometimes

Often

Very often

- Continue

Thank you for your participation. The survey is now finished.

You **MUST click the "Finish" button**, otherwise your participation will be marked as **incomplete** and you will **not be paid**.

After data collection is complete, you will receive information about your result and your payment.

Your answers will be matched with those of randomly selected other participants and evaluated according to the procedures explained earlier.

It is possible that we may not be able to match all responses.

If this happens, unmatched participants will be matched with the average responses of the other study participants and paid accordingly.

- Finish

G Additional games

The following three games were presented after the respondent concluded the public good game phase and before the elicitation of additional demographic and corruption experience questions. The public good phase was always played first, and each additional game was only revealed and described upon its start. Therefore, neither the public good game, nor each additional game, could be influenced by knowledge of what would come next.

G.1 Trust game

The participants played the trust game, once as sender and once as receiver. The order of the two roles was randomised across participants. Only one, randomly selected matching order per pair of participants (see matching protocols in Appendix H), would be valid for payment of the trust game section. All participants were informed of this payment scheme. Moreover, the receiver's choice incentivisation followed the strategy method: an amount sent back would be specified for each possible choice of the sender. The implemented send-back would be the one corresponding to the actual choice of the sender with whom the participant was paired. This payment scheme was transparently disclosed to the participants.

Trust game explanation: `https://youtu.be/_1Oy9cTq66k`

- Continue

This is a checkpoint page, where participants could re-watch the video, click through the slide deck, or continue.

You can watch the video again, click through the slides, or continue the survey by clicking the "Continue" button.

- Continue

- Re-watchable video
- Video stills / slide deck

Sender decision

1You will now take on the role of the **Sender**. Please indicate how much of your 2€, in steps of 0.2€, you would like to send to the other person, if any.

- Continue

Receiver contingent decisions

You are now in the role of the **Receiver**.

Please indicate how much of your 2€ plus the amount you received you would like to send back to the Sender.

Please enter a number in steps of 0.2€.

The choice that applies is the one corresponding to the actual amount sent by the Sender.

2aIf the Sender **sent 0€**, you received 0€. **How much of the 2€ you now have will you send back?** 2bIf the Sender **sent 0.2€**, you received 0.4€. **How much of the 2.4€ you now have will you send back?** 2cIf the Sender **sent 0.4€**, you received 0.8€. **How much of the 2.8€ you now have will you send back?** 2dIf the Sender **sent 0.6€**, you received 1.2€. **How much of the 3.2€ you now have will you send back?** 2eIf the Sender **sent 0.8€**, you received 1.6€. **How much of the 3.6€ you now have will you send back?** 2fIf the Sender **sent 1€**, you received 2€. **How much of the 4€ you now have will you send back?** 2gIf the Sender **sent 1.2€**, you received 2.4€. **How much of the 4.4€ you now have will you send back?** 2hIf the Sender **sent 1.4€**, you received 2.8€. **How much of the 4.8€ you now have will you send back?** 2iIf the Sender **sent 1.6€**, you received 3.2€. **How much of the 5.2€ you now have will you send back?** 2jIf the Sender **sent 1.8€**, you received 3.6€. **How much of the 5.6€ you now have will you send back?** 2kIf the Sender **sent 2€**, you received 4€. **How much of the 6€ you now have will you send back?**

- Continue

G.2 Norm elicitation

Norm elicitation explanation – part 1: `https://youtu.be/5ezYxd1zf8k`

- Continue

This is a checkpoint page, where participants could re-watch the video, click through the slide deck, or continue.

You can watch the video again, click through the slides, or continue the survey by clicking the "Continue" button.

- Continue

- Re-watchable video
- Video stills / slide deck

On the following pages, you will find three statements.
Please indicate how strongly you **agree** or **disagree** with each of them.

- Continue

1**Everyone should cooperate with others in the general interest.**

Strongly disagree

Disagree

Agree

Strongly agree

- Continue

2**Everyone should trust others.**

Strongly disagree

Disagree

Agree

Strongly agree

- Continue

3**No one should betray the trust that others place in them.**

Strongly disagree

Disagree

Agree

Strongly agree

- Continue

Norm elicitation explanation – part 2: `https://youtu.be/5dFPPmF1CBI`

- Continue

This is a checkpoint page, where participants could re-watch the video, click through the slide deck, or continue.

You can watch the video again, click through the slides, or continue the survey by clicking the "Continue" button.

- Continue

- Re-watchable video
- Video stills / slide deck

4**Everyone should cooperate with others in the general interest.** Which answer will be chosen **most frequently** by the **other participants** in this study?

"Strongly disagree" was chosen by most participants

"Disagree" was chosen by most participants

"Agree" was chosen by most participants

"Strongly agree" was chosen by most participants

- Continue

If you guess correctly, you will receive 60 cents.

5**Everyone should trust others.** Which answer will be chosen **most frequently** by the **other participants** in this study?

"Strongly disagree" was chosen by most participants

“Disagree” was chosen by most participants

“Agree” was chosen by most participants

“Strongly agree” was chosen by most participants

- Continue

If you guess correctly, you will receive 60 cents.

6**No one should betray the trust that others place in them.** Which answer will be chosen **most frequently** by the **other participants** in this study?

“Strongly disagree” was chosen by most participants

“Disagree” was chosen by most participants

“Agree” was chosen by most participants

“Strongly agree” was chosen by most participants

- Continue

If you guess correctly, you will receive 60 cents.

G.3 Bomb risk elicitation task (Crosetto and Filippin, 2013)

Bomb risk elicitation task explanation: `https://youtu.be/u0nzcQ8A5BI`

- Continue

This is a checkpoint page, where participants could re-watch the video, click through the slide deck, or continue.

You can watch the video again, click through the slides, or continue the survey by clicking the “Continue” button.

- Continue

- Re-watchable video
- Video stills / slide deck

Your decision

This page displayed the bomb risk elicitation task itself. The task was dynamic: after clicking "Start", one box per second was collected automatically at random. The participant could stop the process at any time. Feedback was enabled, so after stopping, participants could click "Solve" to reveal whether the bomb had been collected. No separate results page was displayed in this app.

Number of boxes collected

Number of boxes remaining

- Start
- Stop
- Solve

- Continue

Part marker shown before the next stage of the study. Depending on prior randomization, participants saw either Part 3, Part 4, or a blank page.

H Asynchronous matching protocol

Public good game groups We maximise the chances of collecting complete responses and minimise procedural complexity by running the experiment asynchronously, i.e. each participant acts as part of a group of 3 but their actions do not require the simultaneous presence of the other group members online, though preserving their strategic and payoff interdependence.

To do so, we assign a sequence number to each participant at the moment in which the experimental conditions are assigned (immediately following the illustration of the basic game). This means that independent of the moment in which the participant will click on the link to reach the survey they received from the panel company, they will be assigned a sequence number the moment they will have learned of the basic procedures of the game and just before being assigned to one of the four experimental conditions.

Three sequentially contiguous participants will be assigned to one of the four experimental conditions. For example, the first three participants to reach the condition assignment stage will be assigned to condition 1; the second three participants will be assigned to condition 2; and so on.

After completing the data collection, we group participants in groups of 3 by assigning each individual a random number q without duplicates. Individual i, non-administrator in experimental condition c, will be assigned to group g as follows:

g = roundup(number of non-administrators in condition c whose q is smaller than or equal to i's, divided by 2).

For administrators in condition c:

g = roundup(number of administrators in condition c whose q is smaller or equal to theirs).

An analogous strategy was followed for role and pair assignment in the trust game, as described below. See preregistration document for a complete description of the procedure.[5]

Administrator selection One of the three group members will be assigned the role of Administrator. This will also be assigned based on the participants' sequence number: as we will match the participants in groups of three with one administrator each, every third participant reaching the condition assignment stage will be assigned the role of Administrator.

Incomplete groups Asynchronous implementation leaves the possibility for the existence of residual participants who cannot be assigned to complete groups within a given experimental condition. This possibility is accounted for in our setting by, at the very end of the study, informing the participants that should it be impossible to assign some participants to complete groups, the behaviours of these participants would be remunerated against the average behaviour of the other participants in the study. This procedure was actually implemented, ensuring that in such case, the average behaviour was computed over the choices made under the same experimental condition that the members of the incomplete groups participated in.

Trust game pairs After completing the data collection, we assign each individual a second random number q without duplicates. Trust-game matching is implemented within experimental condition and separately by public good game role (Group Members and Administrators).

For non-administrators in experimental condition c:

$pair$ = roundup(number of non-administrators in condition c whose q is smaller than or equal to theirs, divided by 2).

For administrators in condition c:

$pair$ = roundup(number of administrators in condition c whose q is smaller than or equal to theirs, divided by 2).

This implies that, within each experimental condition, non-administrators are matched only with non-administrators and administrators only with administrators, pairing them sequentially two by two according to the ordering induced by q. It also implies that trust-game matching is independent of the public good game group assignment itself: public good game groups are constructed using a distinct random number, r, whereas trust-game pairs are constructed using q.

[5]Registered trial AEARCTR-0011134 on `https://www.socialscienceregistry.org/`