Intrinsic Motivation vs. Corruption? Experimental Evidence on the Performance of Officials

Lambsdorff, Johann Graf and Grubiak, Kevin and Werner, Katharina

University of Passau, University of Passau, University of Passau

24 May 2023

Intrinsic Motivation vs. Corruption? Experimental Evidence on the Performance of Officials

Johann Graf Lambsdorff

Kevin Grubiak

Katharina Werner[1]

Abstract

There are conflicting views as to whether corruption or intrinsic motivation plays a greater role in determining the performance of public officials. We run an experiment that incorporates both viewpoints and assess the relative strength and interplay of these respective factors. The design introduces some realism into an everyday exchange between an Estimator (businessperson) and an Auditor (public official) and induces a gray area between intrinsic motivation, extortion and bribery. The Estimator can make a large transfer in the hope of avoiding unfair treatment (extortion) or obtaining an undeserved benefit (bribery). The Auditor may be intrinsically motivated to fulfill her duty or may be corrupted by transfers. We find that intrinsic motivation has a much higher impact on the performance of Auditors than corruption. In a treatment with punishment, Auditors are significantly less likely to accept a large transfer. But punishment fails to bring about favorable welfare effects due to two forces offsetting each other on the individual level. Intrinsic motivation increases for some subjects, supporting the "expressive law" literature, while it decreases for others, supporting the "crowding-out" literature. We infer that punishing officials is an unproblematic tool for fighting corruption, but its effectiveness is called into question. Policies should focus more on preserving officials' intrinsic motivation and worry less about their corruptibility.

JEL classification: K42; D73; C92

Keywords: Bribery; crowding-out; expressive law; extortion; intrinsic motivation

[1] Johann Graf Lambsdorff is professor of economic theory at the University of Passau, Germany. Contact address: Innstrasse 27, D-94032 Passau, jlambsd@uni-passau.de. Kevin Grubiak and Katharina Werner are PostDocs at the University of Passau.

1. Introduction

Corruption, the misuse of public power for private benefit through bribery or extortion, has adverse consequences. It distorts the behavior of public officials, diverts their efforts from their duties and thus undermines their performance to the detriment of economic growth, equity and development. This has been documented in cross-country studies using observational data (for a review see Lambsdorff 2007) as well as in laboratory and field experiments (Serra and Wantchekon 2012). Many regard combatting corruption as a prerequisite for ensuring growth, welfare and development. This spirit has been widely echoed in reports by intergovernmental bodies such as the OECD or the World Bank. The United Nations Convention Against Corruption (United Nations 2014: iii) states: "Corruption is a key element in economic underperformance and a major obstacle to poverty alleviation and development." As Rose-Ackerman and Truex (2012) argue: "Understanding the incentives for corruption and self-dealing is a precondition for making progress on the other challenges facing the world." Without fighting corruption, other efforts to promote welfare are seen as futile. This is in line with a global trend since the 1990s that seeks to increase deterrence through detection and punishment. This development finds support from an economic analysis of crime (Becker and Stigler 1974; Polinsky and Shavell 2000; 2001). A behavioral perspective in the legal literature, known as "expressive law", has also contributed to this view. It argues that punishment and its enforcement deliver a message on social norms that is respected beyond the purely rational calculus.

Opposition has mounted against this focus on corruption and punishment (Kelman 2003; Anechiarico and Jacobs 1996; Gutman 2014; Bandiera et al. 2009). For areas such as the public sector, the biggest challenge may be getting officials to work and exert effort. Whether officials take gifts or bribes is only of secondary importance and may not be crucial to the performance of the public sector. Limiting corruption per se does not induce public officials to perform. Rather this goal should be achieved by management tools, for example those that are aimed at fostering an intrinsic motivation. Deterring corruption might even be at risk of crowding out this intrinsic motivation. It might supply officials with the impression that they are not trusted and that corruption is widespread rather than the exception. At the same time, the detection of corruption requires supervision and control, which takes away discretionary power and autonomy from officials. Such an absence of autonomy might equally impair the officials' intrinsic motivation.

This study tests these two opposing views in a laboratory experiment. The relative importance of intrinsic motivation or corruption can go either way. In doing so, this study goes beyond the existing literature on intrinsic motivation and crowding-out effects. Previous studies have focused on a rather narrow type of intrinsic motivation: The role of ethics, virtue and intrinsic honesty. Given such a narrow focus, practitioners in the area of anticorruption tend to remain unconvinced by the second view. At various events and conversations with such practitioners, during which the authors presented research on the significance of intrinsic motivation with an emphasis on ethics, they were met with skepticism. Many practitioners tend to adhere to the first viewpoint, which emphasizes deterrence. This approach appears to be a

robust response to corruption, employing a hardball strategy, as opposed to a softball approach that relies on ethics. Yet, beyond a narrow focus on ethics, the literature is replete with suggestions of further intrinsic types of motivation, such as the joy of performing a particular task or the commitment to a meaningful purpose. Such a broader recognition of intrinsic motivation might tip the balance in favor of the second view and against the deterrence approach.

We obtain experimental evidence that corruption influences the officials' performance. But this impact falls short of the role played by intrinsic motivation by an order of magnitude. In a treatment where officials are punished, we find that punishment is effective in lowering the acceptance of bribes. But it fails to affect the officials' performance. We do not find evidence for crowding-out effects but instead observe that overall, officials' intrinsic motivation remains unaffected and strong. At the individual level, however, we find that our officials can be classified into two types: Some officials lose their motivation and let performance become more reactive to bribes, which is supportive of the "crowding-out" literature. Yet, for others the role of intrinsic motivation even increases in response to punishment. The latter reaction might be traced to the social norms and duties that are highlighted by punishment, an argument put forward by the "expressive law" literature. It might also act as an individual moral reminder, redirecting attention from pursuing transfers to a commitment towards official duties. Overall, we conclude that management tools aimed at promoting intrinsic motivation should play the main role for the public sector.

2. Literature Review

A long tradition of economic analysis on crime has put a focus on extrinsic incentives, in particular the monetary gains from crime and the risks of detection and punishment. There is now substantial research on the additional role played by intrinsic motivation. It holds that people are characterized by non-monetary and non-standard preferences and might resist crime even if extrinsic incentives render it profitable. This non-standard approach to preferences has also been applied to studies on corruption, the misuse of public power for private benefit through bribery and extortion.

Research has widely investigated the role of intrinsic motivation to resist crime with a focus on the role of ethics. People are regarded to act morally out of virtue or by being reminded of a moral code and the associated social norm. The idea of virtue is that people differ in their level of honesty. Some people are intrinsically honest and resist corruption as a consequence. This link has been emphasized by Gächter and Schulz (2016), who find that participants cheat more in a dice-experiment the higher the level of corruption in their country. Likewise, Cohn et al. (2019) observe a lower willingness to return a lost wallet in countries perceived to be corrupt. Tanner et al. (2022) observe that experimental subjects who are committed to moral principles are less willing to take bribes. Other experimental studies identify methods to increase honesty through moral reminders and framing effects (e.g. Bateson et al. 2006; Hallsworth et al. 2017; Pruckner and Sausgruber 2013).

Still, the influence of virtue and moral reminders is likely to be small. Such reminders might work well in face-to-face interaction or when superiors directly interact with employees.

In a large public office with less personal interaction, ethics might not exert a strong influence. Yet, this study defends the role played by intrinsic motivation. It does so by arguing that intrinsic motivation goes beyond virtue and ethics. The literature is replete with suggestions of further intrinsic types of motivation, such as the joy of performing a particular task, the commitment to a meaningful purpose, the curiosity and eagerness to master a challenge, the pleasure of helping others or contributing to a public good, the desire to comply with social norms, the satisfaction from seeing oneself in a positive light or the preference for self-determination, discretion and autonomy. Thomas (2009: x), for example, suggests four drivers of intrinsic motivation, namely: "a sense of meaningfulness, a sense of choice, a sense of competence, and a sense of progress." Various studies refer to the enjoyment of working on interesting tasks or helping others (Fehr and Falk 2002; Benabou and Tirole 2006). The favorable effect of autonomy on motivation is found in survey work (Carr and Mellitz 2013) and in experiments on workplace environments (Bartling et al. 2012). Its relevance is widely supported in psychology (Patall et al. 2008). These types of intrinsic motivation are not reserved to a few virtuous people. They are also not easily forgotten and thus in need of moral reminders. Instead, they are likely to affect a majority of individuals and retain significance even when they are not deliberately acknowledged. They might thus also work in environments with substantial corruption.

One type of intrinsic motivation relates to autonomy, the absence of control and the freedom to exercise discretion. This is particularly relevant for public offices. Officials may strive to master their bureaucratic tasks or they may enjoy the autonomy in making decisions. In the classical trade-off between rules and discretion, this type of intrinsic motivation may tip the balance in favor of the latter (Spagnolo 2012; Coviello et al. 2018). Some field data demonstrates the strength of this influence. For example, Bandiera et al. (2009: 1284) investigate data from Italian procurement and portray officials in public management as "genuinely motivated to save public money". They infer from their data that this motivation and performance suffer from stricter rules and external controls. In a field experiment with procurement officers in Pakistan, Bandiera et al. (2021) find that "autonomy increases the time devoted to procurement, and this leads to lower prices". Rasul and Rogger (2016) study the impact of autonomy on bureaucratic performance in Nigeria. They find that autonomy correlates positively with project quality, measured by the probability of project completion, while not correlating with increased perceptions of corruption. They also observe management practices of control to backfire. Such practices include collecting indicators of project performance, reviewing these and rewarding bureaucrats for achievements. Rasul, Rogger and Williams (2021) report similar findings for the role of autonomy and incentives among civil servants in Ghana. Bloom et al. (2015) examine differences in school systems and find that students from autonomously operating schools perform better. Our experimental design follows this approach and grants a considerable degree of autonomy to the Auditors. This might contribute to a substantial level of intrinsic motivation, to put effort into fulfilling their duties.

Intrinsic types of motivation may conflict with extrinsic motivation. The introduction of monetary rewards or punishments might crowd out intrinsic motivation (Bénabou and Tirole 2003; 2006). Extrinsic incentives could signal a principal's distrust of an agent's intrinsic motivation or negatively affect the reputational value attributed to a person's actions (Gneezy et al. 2011). Extrinsic incentives also require the measurement of individual achievements and

output. This level of control might limit an individual's autonomy and self-determination, thus impeding intrinsic motivation. A substantial body of experimental evidence supports this crowding-out hypothesis (Deci et al. 1999; Gneezy and Rustichini 2000; Fehr and Falk 2002; Fehr and Rockenbach 2003; Schulze and Frank 2003; Falk and Kosfeld 2006; Dickinson and Villeval 2008; Bowles 2008; Bowles and Polanía-Reyes 2012; Khadavi 2014; Rasul and Rogger 2016; Schildberg-Hörisch and Strassmair 2021).

This risk of a crowding-out effect might also be relevant for areas prone to corruption. Akerlof and Dickens (1982: 318) suggest that imposing severe penalties could undermine individuals' intrinsic motivation to obey the law. Severe penalties may thus induce officials to act as income maximizers, weighing the gains against the risks of detection, and to disregard respect for social norms. Their intrinsic motivation could also suffer if they are frequently checked and feel that they are being treated as suspects.

Increased control could also lead to an adverse spillover effect, such that officials engage in corruption in areas where this is not punished. Galeotti et al. (2021) show such spillovers of crowding-out effects outside the context of corruption. They carry out a quasi-experimental analysis of behavior in public transport. Deboarding passengers were asked by a member of the research team whether they had lost the banknote seemingly picked up from the ground in front of the passenger. Claiming the banknote indicates a passenger's intrinsic dishonesty. Passengers that had previously been checked for fare evasion, that is, whether they had failed to buy a ticket for using public transport, were more likely to claim the bankote. This suggests that deterrence in one domain (fare evasion) may crowd out intrinsic honesty in another domain (not claiming the banknote).

As of yet, experimental evidence on crowding-out in the context of extortion (also called harassment bribery) or bribery (also referred to as collusive bribery) is scarce and largely related to the idea of intrinsic honesty and ethics, without reference to the full set of intrinsic types of motivation. At the same time, it remains uncertain whether crowding-out is the most likely effect to be expected. A huge body of research, mostly in the legal or law and economics literature, often referred to as "expressive law", tends to contradict the crowding-out hypothesis. It suggests that laws express social condemnation and that this induces an intrinsic motivation to abstain from crime (Sunstein 1996). In this way, the law acts as a deterrent by informing people about the behavior expected by society and by providing information on how the majority of people are likely to behave. In a recent contribution, Lane et al. (2023) provide causal evidence on the law's ability to shape the social norms that prevail in a society. Some studies suggest that only those laws that are controlled and enforced obtain an expressive effect. This then implies that deterrence and punishment strengthen rather than weaken people's intrinsic motivation, that is, they crowd in the intrinsic motivation rather than crowding it out. Nosenzo et al. (2022) run an experiment to show that crowding-out occurs when punishment is seen as selfish, that is, when the punisher personally benefits from the punishment. The punishment in this case does not reveal a social norm but only the individual preference of the punisher. When the punishment is imposed for the benefit of others, it expresses the behavior expected from all and induces compliance with this social norm. Prosocial behavior is then promoted by punishment and crowding-in is observed. For a discussion of the more recent literature, see Arlen and Kornhauser (2022).

The role of broad types of intrinsic motivation and whether these are crowded out by sanctions remains to be clarified. This is even more true for extortion and bribery, which are substantially more complex types of crime. Current experiments investigate either extortion or bribery, without recognizing the real-world difficulties in distinguishing between the two. Also, experimental studies on corruption tend to be silent on welfare effects such as the errors in the conduct of an official duty, which are either disregarded or predetermined by design.

The present study contributes to this research gap. It comprehensively assesses the role of intrinsic motivation and its relationship with corruption as a mixture between extortion and bribery. It also examines how intrinsic motivation and corruption contribute to welfare effects in the form of type I and II errors. Further, it determines whether such intrinsic motivation is crowded out by punishment.

The ability of researchers to examine incidents of corruption with observational data is complicated by the hidden nature of crime and by endogeneity issues that make causal inference difficult. To circumvent these challenges, we use a laboratory experiment. Our experimental design models a corrupt exchange with substantial realism to increase external validity. It has become standard practice to model a corrupt exchange as a gift-exchange game where those who make a transfer must hope for reciprocal kindness. Our experimental design follows this practice while adding more realism to it in three respects. First, it is novel in that it motivates subjects intrinsically. It achieves this by endowing both roles (hereafter called the Estimator and the Auditor) with meaningful duties: The Estimator has to guess the number of objects in a picture and the Auditor has to check whether the Estimator's guess is good enough to deserve a prize, a task in which they might easily err. The Auditor enjoys considerable discretion and autonomy in decision making, has to join effort with skill in order to master the given task, and faces varying challenges across rounds. Second, it is novel in that it does not explicitly address the idea of corruption but leaves it up to the subjects to identify this for themselves. Third, it models a corrupt relationship as a gray area between extortion (the Auditor intends to withhold the prize unless given a transfer in the first stage) and bribery (the Estimator makes a transfer in order to get an undeserved prize).

By this, our design allows us to analyze the ambiguous welfare economic consequences of bribery. A type I error (a prize is awarded to an undeserving subject) identifies downside effects of bribery. An Estimator might engage in bribery and obtain the prize that he is not entitled to. At the same time, the design includes the upside of bribery. If the Estimator makes large transfers this reduces the frequency of a type II error (a prize is withheld from a deserving subject). Estimators may be worried about an excessively rigid audit. For example, given circumstances of cumbersome regulation, excessive bureaucracy or market restrictions, some economists tended to emphasize such positive effects of bribe-giving (for a review see Lambsdorff 2007: 2). A large transfer may then compensate for the excessive rigidity. Our design incorporates these two sides of bribery and allows us to infer how punishment affects their relative probabilities.

3. Experimental Design

The design combines standard experiments from the literature where a businessperson must first decide on a transfer and subsequently runs the risk of being cheated by an untrustworthy official (Abbink and Hennig-Schmidt 2006; Abbink et al. 2002; Abbink 2004; Frank and Lambsdorff 2010; Lambsdorff and Frank 2011; Rivas 2013). With respect to duty and autonomy, the design takes inspiration from Gneezy et al. (2019), in which an examiner judges the quality of jokes, using his discretion in favor of a bribing joke writer. Incorporating such a duty has the advantage of vividly capturing the moral costs of rule violation as an externality of the corrupt act, thus avoiding the monetary externalities for lab participants that otherwise add to the complexity of a design. The powerful effect of moral costs on rule compliance has recently been demonstrated in the wide variety of papers on dice-in-the-cup experiments, (Fischbacher and Föllmi-Heusi 2013; Shalvi et al. 2011; Grundmann and Lambsdorff (2017).

In our experiment, two participants interact with one another in the roles of Estimator (representing a businessperson) and Auditor (representing a public official).

[1] The experiment is conducted over 16 rounds, following a stranger-rematching protocol with fixed roles. Three rounds are randomly chosen at the end of the experiment to determine participants' payoffs. The game that participants face in every round consists of three stages as described in detail below.

3.1 The Game

Stage 1: Transfer

At the beginning of every round, the Estimator (he) receives 20 talers and uses it to conduct a dictator game in which he decides how to divide this amount between himself and the Auditor (she). The Estimator can only choose integer values. The transfer can be understood as the granting of an advantage, but it is neutrally described as a transfer in the context of the experiment. The Auditor decides whether to accept the transfer. In the case of rejection, the 20 talers are divided such that the Estimator keeps 12 talers and the Auditor receives 8 talers. On the one hand, this rejection payoff ensures a minimum income of 8 talers for the Auditor. At the same time, it offers her a possibility to reject transfers if she perceives them as an undesired attempt to bribe her.[2]

[1] For the importance of neutral word choice, see Abbink and Hennig-Schmidt (2006).

[2] Note that the Auditor cannot use rejections to financially punish the Estimator, which is what distinguishes our game from an Ultimatum Game (Güth et al. 1982). Conceptually, our game is closest to an 'impunity game', in which responders can voice their disapproval and decide not to accept an offered gift at no financial cost to the dictator (see e.g. Bolton and Zwick 1995; Güth and Kocher 2013: 10). We calibrated stage 1 such that the focal strategy of equalizing payoffs between the roles would require the Estimator to transfer his entire endowment and the Auditor to award the prize regardless of it being deserved or not. Note that other strategies may be equally focal: proposing to split the endowment evenly; proposing just the share the Auditor can secure for herself by rejecting (or just above it to provide financial incentives for acceptance); or proposing nothing (hoping that the Auditor may accept it by mistake). We consider the multiplicity of focal strategies an advantage of our design as it avoids anchoring subjects' decisions in specific directions.

Stage 2: Guess and Award Prize

The second stage of each round is a guessing task. In this task, Estimators and Auditors are shown pictures akin to those shown in figure 1. In every round, the Estimator sees one such picture for 6 seconds and submits a guess regarding the number of objects shown. The Auditor simultaneously sees the same picture for 30 seconds, subsequently receives the Estimator's guess and decides whether this guess is good enough to earn a prize of 20 talers. Neither of the players is informed of the correct number of objects, not even upon completion of a round.

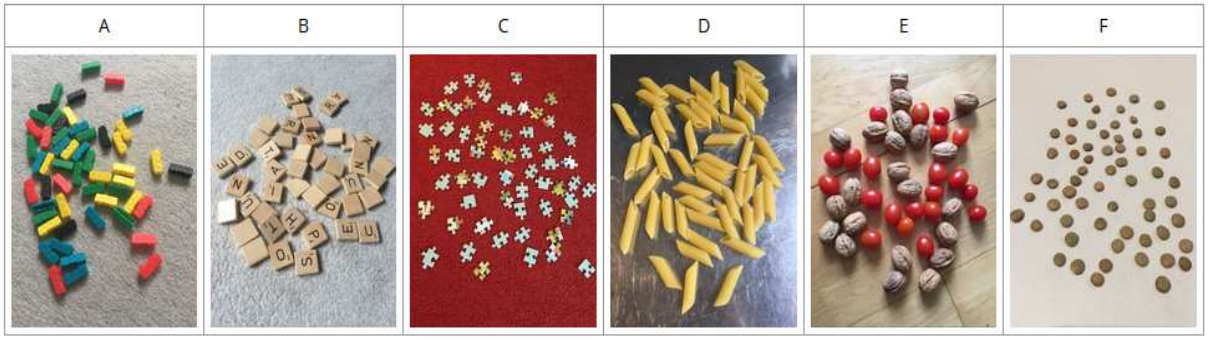


Figure 1: Example Pictures.

Auditors are explicitly instructed to base their award decisions on a prescribed rule: to award the prize if the number of objects is believed to lie within the range of +5/-5 of the true number of objects and not to award the prize otherwise. As such, the Auditor's intrinsic motivation to follow the rule plays a prominent role in this task. At the same time, our task lets Auditors gather their own (possibly erroneous or biased) evidence regarding the correct solution. This design choice endows Auditors with discretionary power, combined with uncertainty about whether their choice has been influenced by the transfer. Auditors may accept a large transfer and award the prize, forming a biased belief that the guess was good enough. Likewise, they might be dissatisfied with a low transfer and withhold the prize, justifying this decision with a biased belief that the guess was not good enough. Thus, corrupt reciprocity need not involve a conscious violation of duties. This introduces a degree of realism into the design, in line with widespread evidence of biased perceptions among those guilty of corruption. At the same time, it should guarantee a large influence of transfers on award decisions.

By combining the two stages, our design allows us to examine the relationship between the Auditor's corruptibility (as induced by stage 1 transfers) and her intrinsic motivation (devoting effort into fulfilling her duty). The strength of our experimental design is that it provides scope for both extortion and bribery. The Estimator may feel extorted insofar as he fears an unfair withholding of the prize without a transfer, i.e., a type II error. Bribery, on the other hand, occurs in the case of a large transfer and an overly generous examination, which implies a type I error. This gray area between extortion and bribery reflects the common difficulty of distinguishing between the two in the real world. For example, businesspeople often claim that they were extorted and only paid a bribe out of concern for unfairness. In most cases, it remains unclear whether they were actually trying to gain an undeserved advantage. At the same time, in the case of officials, it is often impossible to determine whether they extorted a payment or accepted it rather reluctantly.

Stage 3: Review

In the third and final stage of each round, a review may take place. This review is unbiased, i.e., not influenced by a transfer, but not free of error. For this review, we used data from randomly selected subjects who had participated in a pilot experiment and whose only task had been to guess the number of objects in the same pictures, in exchange for an accuracy-dependent financial reward. These subjects (referred to as Inspectors) had faced the same time constraint as Auditors in the present study, but their task was not embedded in a strategic context that would open the door to corruptibility.[3]

A review retrieves information from the previous two stages of the current round and displays it to both players. Specifically, the players are informed

i) whether a transfer greater 8 was proposed and whether such a transfer was accepted or not

ii) whether an Inspector would have awarded the prize or not

Subjects are informed in advance that the Inspector's award decision is objectively correct only about 80% of the time. This is to avoid naïve assumptions of an error-free review process. Besides modeling the review process more realistically, incorporating noise into the review process also minimizes opportunities for Auditors to assess the correctness of their decisions across rounds, thereby preserving their room for discretion.

In any given round, a review takes place with a probability of 25%. Apart from this, both the Estimator and the Auditor can request a review at a personal cost of 1 taler, thereby enforcing a review. No information is communicated to players as to whether a review was requested by the other player or implemented by the computer.

In the baseline version of our game, the outcomes of a review are merely informative. In our treatment, which we describe in more detail in Section 3.3, we test the effectiveness of a popular method of containing corruption in the field: punishing the official for accepting a bribe.

3.2 Behavioral predictions for the baseline

Auditor predictions

The decision whether to award the prize has no monetary consequences for the Auditor in the baseline. This suggests two opposing views on behavior. The first one focuses on corruption. An Auditor is motivated by positive and negative reciprocity in response to the transfer. She will form beliefs about a fair transfer and award the prize for all transfers above this level. The Estimator's guess would play no role. The second view focuses on the Estimator's guess alone and disregards issues of reciprocity. The Auditor is intrinsically motivated to fulfill the assigned duty. She devotes effort, scrutinizes the picture and comes to an unbiased decision on the award of the prize. In the following, we combine these two views into a single model. The extreme views can then be derived by choosing respective parameter values.

[3] The guess of such a randomly selected Inspector was automatically compared to that of the Estimator to determine whether the Inspector would have awarded the prize (deviation < 6) or not (deviation >5).

Following Rabin (1993) and Dufwenberg and Kirchsteiger (2004), the behavior of the Auditor i can be derived from her utility function in round t (neglecting subscripts i and t):

$$U = T - c \cdot \tau + \lambda(\dots).$$

Here, the monetary gain of the Auditor is determined by the received transfer T and the intrinsic motivation c, determined here as the marginal cost of violating the rule. The total cost of rule violation $c \cdot \tau$ increases linearly with the tolerance τ that is required for an award of the prize. If the Estimator's guess differs from the Auditor's solution by more than 5, say 8, then a tolerance of $\tau = 3$ is required to award the prize. The larger τ, the more the Auditor would have to bend the imposed rule to tolerate an excessive deviation.

With λ, we denote the Auditor's propensity for reciprocity. For this, the Auditor forms a belief about the Estimator's kindness expressed in the dictator game. In the dictator game, the Estimator chooses from all available integer values between 0 and 20 talers. The median value of this range is 10 talers, called the equitable allocation. A transfer of 10 talers would thus be neutral. If the transfer exceeds 10 talers, this is perceived as an act of kindness. This induces a tendency towards a reciprocal action in favor of the Estimator, i.e. an award of the 20 talers prize.

A preference for reciprocity is determined by the product of perceived kindness and the reciprocal gain from awarding the prize (Rabin 1993). As such, the following holds after an award of the prize:

$$U_{Award} = T - c \cdot \tau + \lambda 20(T - 10)$$

Not awarding the prize would result in $U_0 = T$. The prize is awarded if $U_{Award} \geq U_0$. In this case, the disutility from violating the rule $-c \cdot \tau$ falls short of the reciprocal gains, which consists of the 20 talers award multiplied with the kindness $(T - 10)$ and the propensity to reciprocate λ. This is met, if $\tau \leq \frac{\lambda}{c} 20(T - 10)$. We define $\bar{\tau} = \frac{\lambda}{c} 20(T - 10)$ as the maximum deviation that the Auditor is willing to tolerate conditional on the transfer. The Auditor will award the prize if the Estimator's guess deviates from the Auditor's estimate by $5 + \bar{\tau}$ or less. This allows us to state the first hypothesis about behavior in the baseline game. We hypothesize the award probability to depend negatively on the deviation of the Estimator's guess from the true value.

> **H1 [intrinsic motivation]:** The deviation of the Estimator's guess from the true value has a significantly negative impact on the Auditor's decision to award the prize.

This first hypothesis is due to the Auditors' intrinsic motivation to follow their duty and to be attentive to the Estimator's guess. The maximum tolerance $\bar{\tau}$ is a linearly increasing function of the transfer that was offered by the Estimator. This allows us to set up our second hypothesis.

> **H2 [corruption]:** Offered transfers have a significantly positive impact on the Auditor's decision to award the prize.

The maximum tolerance $\bar{\tau}$ is also linearly proportional to the coefficient of corruption (the Auditor's propensity for reciprocity λ) and inversely proportional to the coefficient of intrinsic

motivation (the marginal cost of rule violation c). For T=10, which is regarded neither as kind nor as unkind, we get $\bar{\tau} = 0$, i.e. zero tolerance. Transfers above 10 induce reciprocity and hence, $\bar{\tau} > 0$. Assuming, for example, $\frac{\lambda}{c} = 0.05$, a transfer of $T = 20$ would buy a maximum tolerance of $\bar{\tau} = 10$. This means that a deviation of the Estimator's guess from the Auditor's solution of up to (10 + 5 =) 15 would be tolerated. At a transfer of $T = 4$, negative reciprocity would be induced. Instead of zero tolerance, an Auditor will be excessively intolerant with $\bar{\tau} = -6$, in which case she will never award the prize, not even if the guess was perfectly correct.

Which of the two variables, corruption or intrinsic motivation, has a stronger influence? An answer to such a question depends heavily on the distribution of the respective variables. We thus specify the question: Does an increase of a transfer by one standard deviation (SD) influence the award probability more strongly than a decrease of the deviation by one SD? In the model, this would of result if $\frac{\lambda}{c}$ were large. Based on the prevailing focus in the literature, we hypothesize:

> **H3 [corruption trumps intrinsic motivation]:** Transfers have a significantly higher influence on the Auditor's decision to award the prize than the deviation from the true value.

For further details on the model, including a calibration and the derivation of the optimum behavior, see Appendix B.

3.3 Treatment-specific hypotheses

Our treatment differs from the baseline with respect to the consequences associated with specific outcomes of the review process. In the baseline, a review is purely informative and does not have any deterrent effect. We use this baseline as a benchmark against which to compare the deterrent and welfare economic consequences of a popular intervention to curb corruption as summarized in table 1 below.

Treatment	**Consequence**
Baseline	Feedback only
Treatment: Punishing acceptance of a transfer > 8 Taler	8 Taler deducted from Auditor

Table 1: Treatments and consequences of a review

Table 1 depicts the consequences of a review. In the treatment, Auditors are punished by deducting 8 talers from their income if the review identifies that a transfer > 8 was accepted.[4]

[4] We did not employ the equitable payoff T=10 as a threshold for punishment but a lower level of 8. This was chosen because this is the default that is given to the Auditor in case the offered transfer is rejected.

If no review is carried out, no consequences are imposed on any player (regardless of their behavior).[5]

Punishing the Auditor renders the acceptance of a transfer greater than 8 less attractive. Due to the 25% random probability of a review, accepting a transfer imposes an expected cost of $8 \cdot 25\% = 2$. Transfers are less efficient as a tool for inducing reciprocity. In order for transfers to induce an equal level of reciprocity, transfers would have to increase by 2. Estimators who understand this and still aim at bribing the Auditor will therefore increase their transfers. As preregistered, we thus hypothesize that overall transfers will increase. Given that transfers are less efficient in inducing reciprocity and that they are costly, the increase in transfers is likely to be modest, at best. This allows us to hypothesize that Auditors will more often reject the offered transfer. Overall, we thus hypothesize:

> **H4 [Punishment of accepting transfers]:** In the treatment, transfers will increase and they are more often rejected.

Given the reduced reciprocal incentives for Auditors, one might infer that the probability to award the prize will go down. In such a case, the probability of a type II error will increase and that of type I will decrease. Yet, in line with the crowding-out literature, the effect might just as well go in the opposite direction. This is so because the punishment in the treatment may act against intrinsic motivation, potentially inducing Auditors to be lenient and thus to award the prize with a higher probability. While the effect on the award probability is overall unclear, we might nonetheless find evidence in favor of a crowding-out effect. A punishment might distract the Auditor from her duty and thereby reduce the intrinsic motivation.[6] This could happen for a number of reasons. First, Auditors may rationalize the punishment as an incentive against accepting bribes. This idea might create a self-fulfilling expectation. "If everyone else is taking bribes," the thought may go, "I should be no different." Second, the punishment could be seen as a signal of mistrust that takes away some of the Auditors' autonomy. Third, the punishment reduces the Auditor's income, which is then substantially below that of the Estimator. In response, the Auditor might shift attention to the size of transfers, seeking to be compensated for the loss of income. Building on hypothesis 1, we hypothesize a lower influence of the deviation on the Auditor's award decision in the treatment.

[5] Note that the neutral description of transfers is common to both the baseline game and the treatment, i.e. in neither of them do we explicitly convey moral prescriptions regarding the transfer. In the treatment, the punishment is described in a similarly neutral fashion as a monetary deduction based on the outcome of the review process.

[6] In addition to a crowding-out effect, also another effect might induce a higher award probability: For an Estimator who was withheld the prize, there exists a behaviorally plausible motivation for a review: negative reciprocity. The Estimator may be willing to bear the cost of 1 taler in order to harm the Auditor. If the Auditor anticipates this, she worries about the risks of withholding a prize. Once she accepts a transfer larger than 8, she lives at the mercy of the Estimator and is likely to award the prize. Estimators who understand this will try to lock the Auditor in through the payment of such a transfer. We initially favored this effect in our preregistration, but eventually decided to discuss it more thoroughly in a companion publication.

H5 [crowding-out]: The deviation has a weaker impact on the Auditor's decision to award the prize in the treatment compared to the baseline.

3.4 Experimental procedures

The experiment was programmed using oTree (Chen et al. 2016) and conducted at two experimental laboratories in Germany: the PAULA Laboratory at the University of Passau and the KLAB Laboratory at the University of Kassel. We recruited a similar number of participants at both locations using the recruitment software Orsee (Greiner 2015). In total, we had 250 participants take part in one of 12 sessions. The experiment formed part of a broader research agenda pre-registered at AsPredicted.org as 'Between reciprocity and responsibility – anti-corruption experiments' (aspredicted.org, #80975). The files and the original German instructions for running the experiments are available upon request.

English translations of all written and oral instructions can be found in Appendix A. Upon arrival, subjects were randomly seated in the laboratory and publicly instructed (A.1) about the purpose of the experiment, its expected length, the conversion of the experimental currency unit taler into euros, the rules of behavior and (standard) payment and blindness procedures. Subjects then privately read the experiment's core instructions (A.2), which were identical and independent of their later-assigned roles, followed by comprehension questions (A.3) that had to be solved correctly to proceed. Subjects received 1 taler for each comprehension question they answered correctly in the first attempt. The experiment concluded with a role-specific questionnaire (A.4) that asked subjects to elaborate on their decisions in the experiment and involved questions to assess (in an incentivised way) the absolute and relative accuracy of their decisions, and to provide demographic information.

The experiment lasted about 60 minutes. Final payoffs were converted from talers into euros and paid out to the participants at the end of the experiment by a third person who did not know the experiment. The conversion rate was constant across treatments and amounted to 1 taler = €0.20. In both cities, the variable payoff amounted to 9.30€. According to different laboratory standards, the fixed payment for completing the experiment differed (€3.50 in Passau and €5 in Kassel). The average payment in the experiment was thus €12.80 in Passau and €14.30 in Kassel.

4. Results

4.1 Sample overview and summary statistics

Overall, 250 subjects participated in the experiment, 124 (62 pairs) in the baseline and 126 (63 pairs) in the treatment. With 16 rounds, we obtain 992 observations from the 62 Estimators and 992 from the 62 Auditors in the baseline. In the treatment, we obtain 1008 observations from the 63 Estimators and Auditors, respectively.

There are large variations in the deviations between the Estimator's guesses and the true values, as shown in figure 2. The standard deviation amounts to 8.5. A perfect guess with a deviation of 0 occurs with a probability of 3%. Deviations of 1, 2, 3 or 4 are more likely with

a probability of around 8% each. They are more likely than a deviation of 0 because they can be caused by guesses both above and below the true value. Deviations of 5 or less, for which Auditors should award the prize, occur with 40% probability. Deviations above 5 become less likely as the deviation increases. This suggests that Estimators invest effort in attempts to guess well. This may be indicative of their desire to be good at guessing, or because they think that Auditors are intrinsically motivated to scrutinize the quality of their guesses.

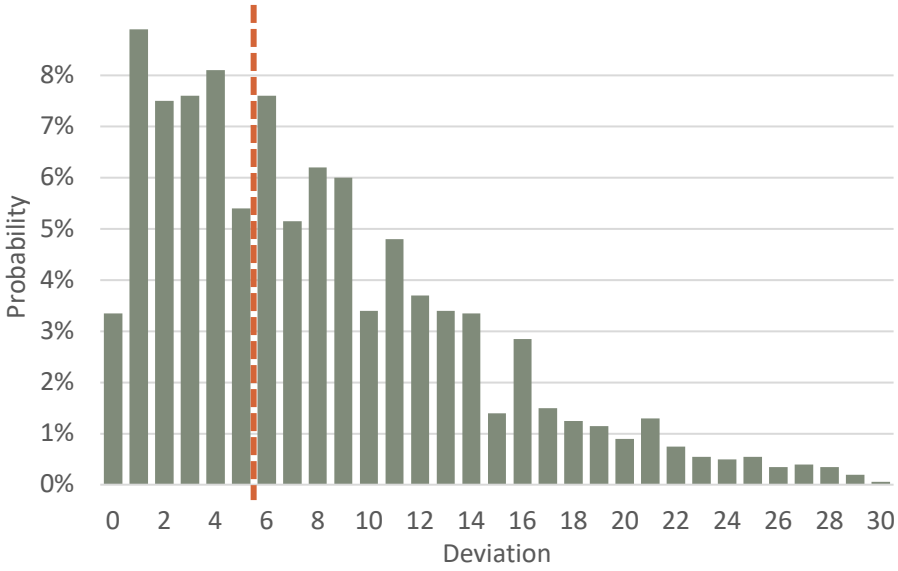


Figure 2: Histogram for deviations (baseline and treatment)

Offered transfers exhibit values spanning the entire available range between the minimum of 0 to the maximum of 20, as shown in figure 3. The value 8 stands out, as this is the level that the Auditor can secure by rejecting the offered transfer (indicated by red color). An equal split of 10 is offered with the highest probability of almost 20%. A standard deviation of 3.9 corresponds to substantial variation in the offered transfers. 57% of the transfers are above the level of 8. Unsurprisingly, transfers above 8 are mostly accepted while those below 8 are almost always rejected. As a consequence, offers are accepted 59% of the time.

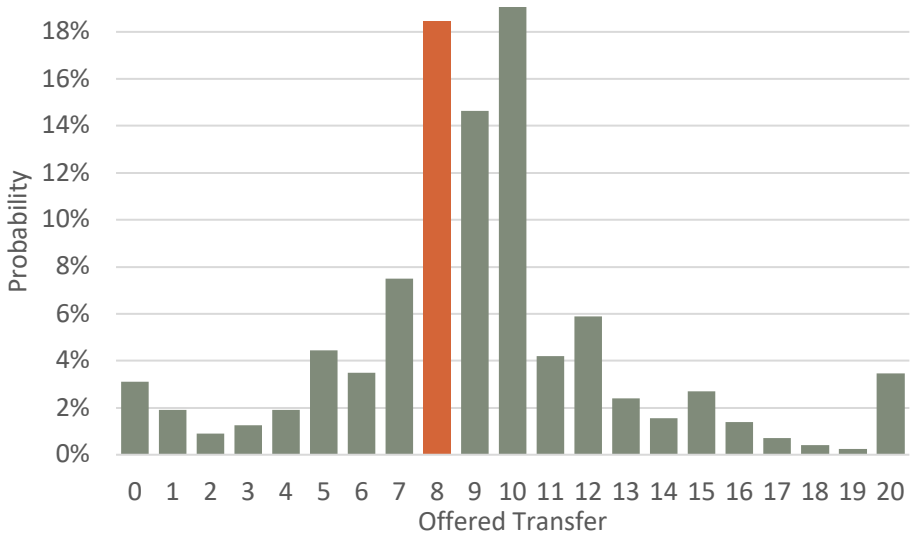


Figure 3: Histogram for offered transfers (baseline and treatment)

Estimators might consist of two types: The untalented who are bad at guessing and who seek to improve their fortunes by making a large transfer, and those with sophisticated guessing skills who feel no need to bribe. If this were true, we would expect a positive correlation between deviation and transfer. But the correlation coefficient amounts to almost zero (0.04). We observe that talented and untalented Estimators have an equal propensity to make large transfers. This implies that, when forming expectations regarding the quality of a guess, Auditors cannot infer that a large transfer signals bad quality. For our subsequent regressions, this implies that the influences of deviations and transfers are most likely orthogonal. On average, Auditors award the prize with 40% probability. By coincidence, this is equal to the probability with which Estimators would deserve a prize. Unsurprisingly, the match between submitting a good guess and getting a prize is far from perfect. This induces errors of type I and II, for which we will show data below.

4.2 Intrinsic motivation and corruption in the baseline

We defined intrinsic motivation as the effort to follow the duty and award the prize based on the quality of the Estimator's guess. Figure 4 presents summary statistics on this by illustrating the probability of Auditors awarding the prize if they should have awarded it (deviation <6, in the back) or should not have awarded it (deviation >5, in the front). The bars on the left ("Baseline") represent the mean award probability from our baseline. As can be seen, intrinsic motivation has a substantial impact. When the Estimator's guess is not good enough, the probability of awarding the prize amounts to only 22%. In this case, an error of type I would occur, the erroneous award of a prize to an undeserving Estimator. When the guess is good, the award probability increases to 61%. An error of type II occurs when the guess is good enough but the prize is withheld, which arises with the counter probability of 39%. The two bars show an increase in the probability from 22% to 61%, revealing a strong willingness (absent any monetary incentives) on the part of Auditors to devote effort into scrutinizing the picture and to decide in line with their duty.

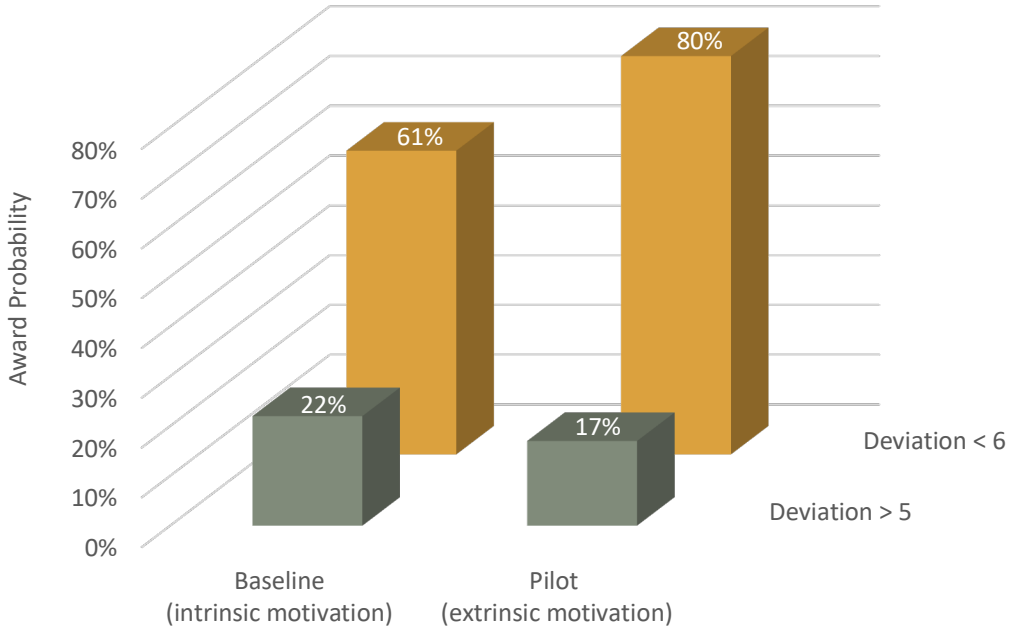


Figure 4: Probability of awarding the prize by type of motivation and deviation

We can determine the strength of the intrinsic motivation relative to a first-best case where Auditors are extrinsically motivated. The bars on the right side ("Pilot") serve as such a benchmark. They are constructed using data from the previously mentioned pilot experiment, in which we gave Auditors a reward for accuracy. The Auditors saw an Estimator's guess. They were asked to provide their own estimate, rather than make an award decision, and received a reward if their estimate was close to the true value. This benchmark reflects how Auditors would decide when incentivized to judge according to the best of their knowledge and in the absence of transfers and any other-regarding preferences. If the deviation of the Estimator's guess from the Auditor's estimate is less than or equal to 5, we treat the data as if the Auditor had awarded the prize; if the deviation is greater than 5, we treat the data as if the Auditor had not awarded the prize. As shown on the right-hand side of figure 4, around 80% of Auditors would award the prize if they were required to do so, and 17% of Auditors would award the prize if they were not required to do so. The perfect score of 100% award probability for small deviations and 0% for large ones is unachievable due to the inability of Auditors to estimate with perfection. The benchmark provides a reasonable estimate that takes this inability into account. A comparison of the benchmark with the performance in our baseline shows that intrinsically motivated Auditors do not perform as well as those in the first-best case, but their performance does not lag far behind. Overall, this supports hypothesis 1.

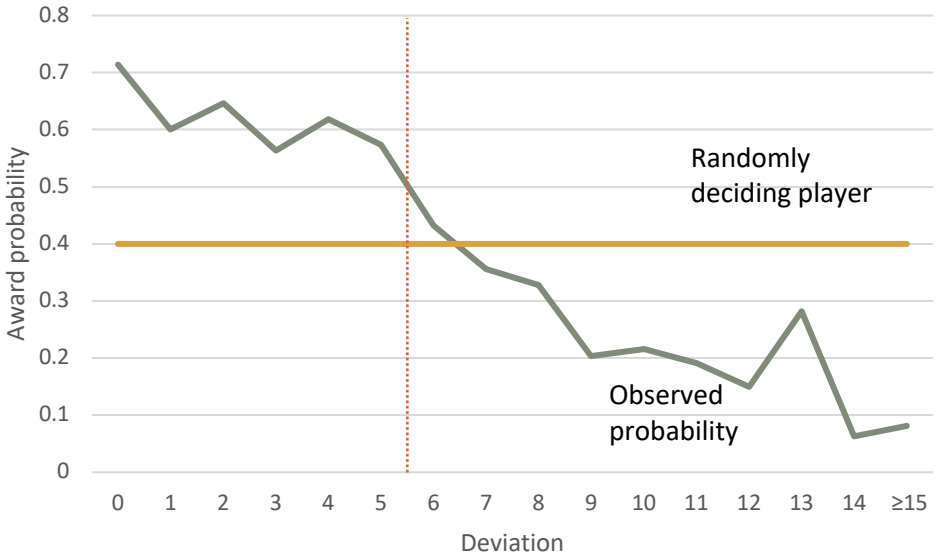


Figure 5: Probability of awarding the prize conditional on the deviation

Figure 5 plots the mean award probability as a function of the deviation. The curve is downward-sloping with values around 60% for small deviations and below 20% for large deviations. A vertical line between 5 and 6 indicates the threshold that Auditors should dutifully consider. Unsurprisingly, the difference between a deviation of 5 or 6 is difficult for Auditors to assess and the award probabilities do not drop as drastically as they should. A deviation of 6 is associated with an award probability of 43%, still above the average of 40%. Auditors, contrary to their duty, retain a high likelihood to reward these guesses. This finding might also be influenced by an anchoring effect. The Estimator's guess might distort the Auditor's judgment. For example, if the Auditor counted 45 objects and sees that the Estimator guessed a higher number, she might update her number to 46. We determined the magnitude of this anchoring effect from our pilot experiment and observed that it amounts roughly to 1. This might explain why a deviation of 6 still induces an above-average award probability.

Imagine a fictitious benchmark player who acts without any intrinsic motivation, devoting no effort into scrutinizing the pictures, and who has no interest in deciding correctly on the award of the prize. Such a benchmark player might award the prize with the average 40% probability, which incidentally is also the average probability with which the prize should have been awarded. Such a player would make substantial errors, both of type I and type II. The intrinsic motivation of our Auditors in the baseline makes them perform much better than such a benchmark player. To the left of the dashed line, the Auditor should always award the prize. The award probability is well above the benchmark, i.e. our Auditors make fewer type II errors than the benchmark player. To the right of the dashed line, the Auditor should never award the prize. The award probability is overall below the benchmark, i.e. our Auditors make fewer type I errors than the benchmark player. Overall, figure 5 reveals that the intrinsic motivation of the Auditors in the baseline is substantial and leads to a performance that is clearly better than that of the fictitious benchmark player.

To test H1 quantitatively, we regress the Auditor‘s decision to award the prize on the deviation, see table 2. We control for the offered transfer. To allow for a subsequent comparison of the magnitude of these two competing forces, we standardize both variables such that both

have a mean of 0 and a standard deviation of 1. Since the dependent variable (the decision to award the prize or not) is binary, we use logit regressions. We cluster our standard errors at the individual level to account for the fact that we have multiple observations per participant. In addition to a first regression (1), we add control variables in a second regression (2). These include some subject characteristics (age and gender) and the location of the sessions. We also include round dummies to account for the possibility that the effort to scrutinize the pictures or experience in counting objects change over time, or that some pictures are more challenging than others. As can be seen, these controls are immaterial to our results.

	(1) Prize awarded	(2) Prize awarded
Deviation	-1.44***	-1.45***
(standarized)	(0.18)	(0.19)
Offered Transfer	0.41***	0.41***
(standardized)	(0.09)	(0.09)
Constant	-0.79***	-1.90***
	(0.12)	(0.66)
Controls	No	Yes
Observations	992	976
Pseudo R^2	0.175	0.189

Logit regressions with clustered standard errors in parentheses
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table 2: Logit regressions of the award decision in the baseline on the (standardized) deviation and the (standardized) transfer

The coefficient of -1.44 for the deviation implies that the deviation has a highly significant negative impact on the decision to award the prize, thus supporting H1. Imagine an increase in the deviation from 4 to 12 (roughly by one SD). Such an increase would reduce the odds ratio of being awarded the prize, for example by slighly more than from 60% (odds ratio $0.6/(1-0.6) = 1.5$) to 20% (odds ratio $0.2/(1-0.2) = 0.25$).

> **R1 [intrinsic motivation]:** The deviation has a significantly negative impact on the Auditor's decision to award the prize.

A second force in the decision to award the prize, which could overshadow the impact of the intrinsic motivation, is corruption. The offer of a generous transfer might tempt Auditors to neglect their duty and award the prize even if the Estimator's guess was not good enough. A high offered transfer is thus likely to induce a type I error. A low offered transfer could have the opposite effect: A transfer below 10, the equal split, or even below 8 is likely to be regarded as unkind. Auditors who are offered such a low transfer might therefore withhold the prize even if the Estimator deserves it, thus committing a type II error.

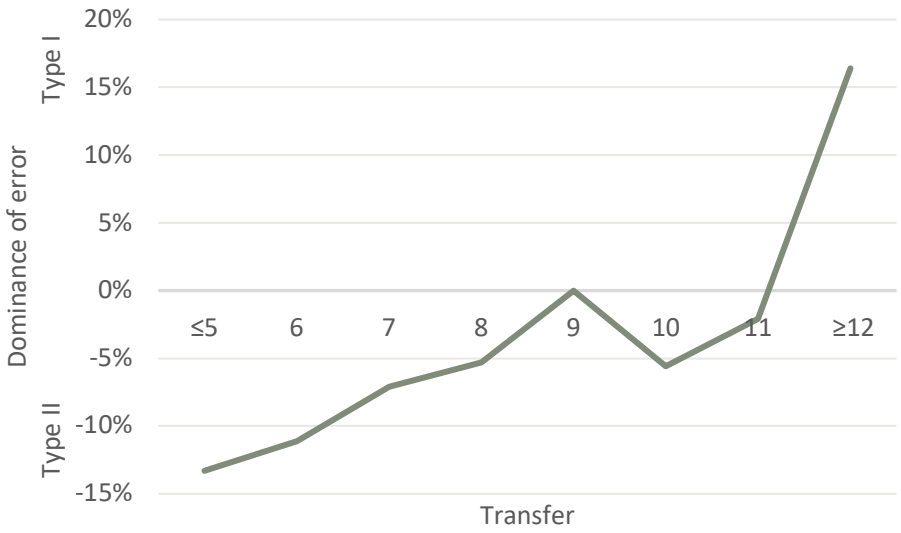


Figure 6: Probability of type II and type I errors based on the offered transfer

Figure 6 plots this relationship. We determine the probability of a type I error relative to a type II error. If, for example, Auditors erred in 10% of all decisions by wrongfully awarding the prize and in 25% of cases by wrongfully withholding the prize, we obtain a -15% dominance of errors. This variable, determined across all available observations in the baseline, is plotted on the y-axis, conditional on the offered transfers on the x-axis. There is a clearly positive slope, indicating corrupt reciprocity. For transfers below 8, there is a predominance of type II errors. On average, Auditors react to small transfers by withholding the prize. Only for transfers greater than 11 is there a clear predominance of type I errors. Auditors positively reciprocate large transfers by committing type I errors.

To quantify the effect of reciprocity on the probability of awarding the prize and to test H2, we return to the regressions in table 2. The coefficient of 0.41 for the transfer implies that the transfer does indeed have a highly significant positive effect on the decision to award the prize, thus supporting H2. Consider an increase in the transfer from 8 to 12 (roughly by one SD). Such an increase would raise the odds ratio of being awarded the prize, for example from 20% (odds ratio $0.2/(1-0.2) = 0.25$) to slightly less than 40% (odds ratio $0.4/(1-0.4) = 0.67$).

> **R2 [corruption]:** The offered transfer has a significantly positive impact on the Auditor's decision to award the prize.

In H3, we hypothesized that Auditors would give more weight to the offered transfer than to the deviation from the true value when deciding whether or not to award the prize.

Figure 7 illustrates the Auditors' probability of awarding the prize, as in figure 4. The wrongful award despite a large deviation (deviation > 5) is shown in the front. The rightful award for small deviations (deviation < 6) is shown in the back. The bars on the left show the mean award probability when the offered transfer was less than or equal to 8, i.e. the minimum transfer an Auditor would receive in case of rejection. The bars on the right side display the mean award probability if the offered transfer was greater than 8.

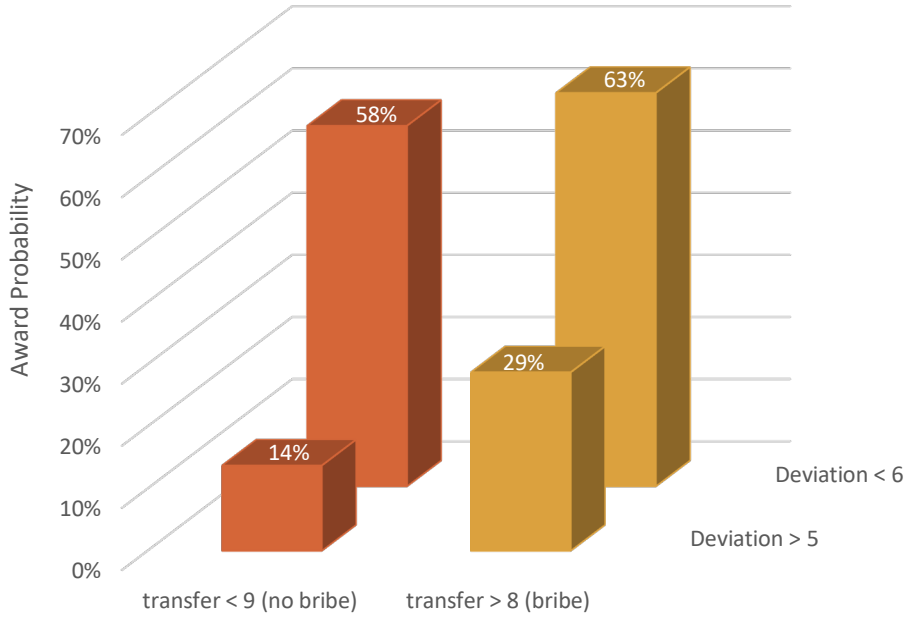


Figure 7: Probability of awarding the prize by transfer and deviation

A comparison of the values shows that the deviation has a much greater influence on the award probability than the transfer. If the deviation is small rather than large, the award probability increases from 14% to 58% and likewise from 29% to 63%, roughly by 40 percentage points. On the other hand, if the offered transfer is high rather than low, the award probability increases from 14% to 29% and likewise from 58% to 63%, on average by 10 percentage points. Intrinsic motivation has a higher impact than corruption by an order of magnitude, which contradicts our hypothesis 3. We can test this hypothesis quantitatively on the basis of the regressions in table 2. A Wald test for the equality of coefficients confirms that the coefficients are significantly different from one another ($p<0.001$).

> **R3 [intrinsic motivation trumps corruption]:** The deviation from the true value has a significantly higher influence on the decision to award the prize than the transfer.

4.3 Punishment and Crowding-out

Figure 8 depicts mean outcomes for the baseline and the treatment. In the treatment, transfers did not increase, as hypothesized, but even decreased slightly relative to the baseline from 9.3 (46% of the maximum of 20) to 8.9 (45%). The difference itself is not significant ($p=0.66$, Wilcoxon signed-rank test). In line with H4 we observe that Auditors less often accept transfers, namely in 49% of the cases, which is significantly below the acceptance rate of 70% in the baseline ($p<0.001$). The punishment is thus capable of reducing the acceptance of transfers, which could be understood as “bribes” in this context.

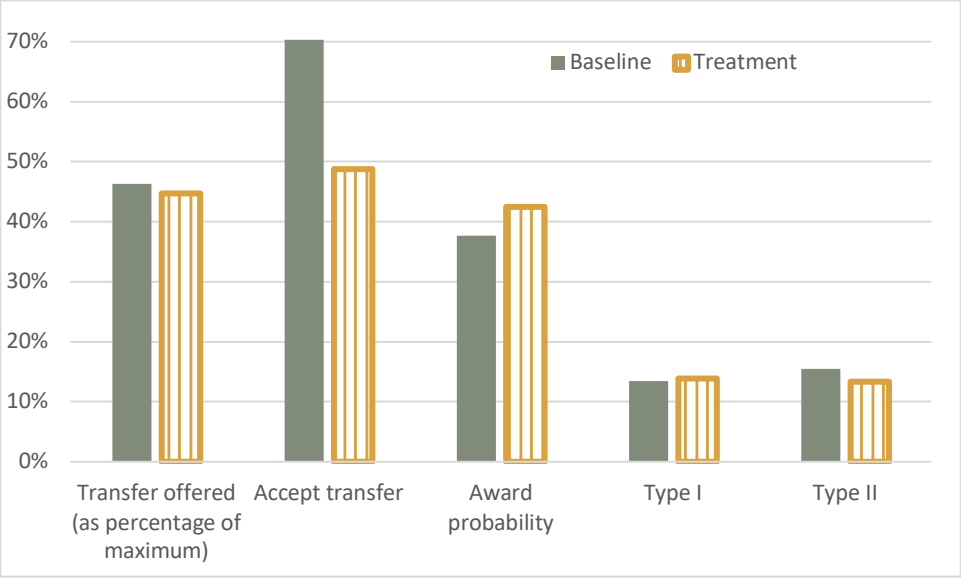


Figure 8: Mean data (baseline and treatment)

We observe a low acceptance in particular for transfers below 10. Auditors might have formed rational expectations and correctly assessed the risk of detection against a secure income of 8. Another reason for the significant finding could relate to the expressive function of the treatment. The punishment for accepting transfers may have acted as a moral reminder or revealed a social norm that finds the taking of bribes despicable.

> **R4 [Punishment of accepting transfers]:** In the treatment, transfers do not increase and are significantly more often rejected.

Regarding the award probability, we observe that it slightly increases from 38% to 42%. The difference is not statistically significant ($p=0.19$). The observed increase may suggest a crowding-out effect. Yet, the evidence remains weak. The increased award probability translates to a similar finding regarding a type I error. The probability increased from 13% to 14%, which is not significant ($p=0.73$). Likewise, the probability of a type II error insignificantly decreases from 16% to 13% ($p=0.20$).

We hypothesized in H5 that Auditors would be less intrinsically motivated in the treatment than in the baseline. This can be directly tested using regression analysis. Table 3 tests the influence of the deviation and the transfer in the treatment and allows for a comparison of the coefficients with those from table 2. The influence of transfers is rather small with a coefficient of 0.25. This is clearly below the influence that was found in the baseline with 0.41 (see table 2). This finding is in line with data shown in figure 7. The punishment puts a limit on the influence a transfer can buy.

If there were a crowding-out effect, we would expect the coefficient of the standardized deviation to be smaller in absolute value in the treatment than in the baseline. With a value of -1.43, we obtain a value that is largely identical to the one from table 2, where we obtained a value of -1.45. Unsurprisingly, a Wald test for equality of the two coefficients does not find a significant difference ($\chi^2=0.01$). We thus cannot support H5.

	(1) Prize awarded	(2) Prize awarded
Deviation	-1.40***	-1.43***
(standarized)	(0.19)	(0.19)
Offered Transfer	0.22**	0.25**
(standardized)	(0.10)	(0.10)
Constant	-0.52***	-0.86
	(0.13)	(0.68)
Controls	No	Yes
Observations	1008	1008
Pseudo R^2	0.160	0.175

Logit regressions with clustered standard errors in parentheses
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table 3: Logit regressions of the award decision in the treatment on the (standardized) deviation and the (standardized) transfer

R5 [no crowding-out]: The deviation from the correct value does not have a weaker impact on the probability of awarding the prize in the treatment compared to the baseline.

The absence of a crowding-out effect comes by surprise. One explanation might be that the Auditors do not perceive the mere threat of punishment as a loss of autonomy or as a signal of mistrust. They do so only after having experienced an actual punishment. Auditors might then engage with a high level of intrinsic motivation until an actual punishment takes place. As long as an Auditor is not punished, her experience does not differ from that in the baseline. This could cause the two coefficients to be similar.

We test this explanation by splitting our data into two groups and running the regressions separately for each, as shown in table 4. Column (4) refers to the treatment group, which includes only those 183 observations where the Auditor was punished in the preceding round. Likewise, column (3) includes the 763 observations where the Auditor did not experience punishment in the preceding round. We replicate a similar split for the baseline. Column (2) includes the 236 observations where a review identified the same type of misconduct that induces a punishment in the treatment: A transfer greater than 8 had been offered and accepted and a review had been conducted. In the baseline, this misconduct is reported without invoking a punishment. Column (1) includes the 679 observations where no misconduct took place.

If experienced punishment crowds out intrinsic motivation, the coefficient of the deviation should be smaller in absolute terms in column (4) than in column (3). This effect should not be present in the baseline. Indeed, as can be seen in table 4, the respective coefficient does not differ between columns (1) and (2). Contrary to our expectations, in the treatment the coefficient in column (4) is not smaller but larger in absolute terms than the one in column (3). The unexpected difference is even significant at the 10%-level (Wald-test). This evidence does not support a crowding-out effect. Rather, it suggests that the punishment serves as a signal to

the Auditor that the behavior detected was wrong. This is indicative of an influence that runs counter to crowding-out. This influence could be described by reference to the literature on expressive law. Auditors might see the treatment as a moral reminder or as an expression of a social norm to refrain from bribery and to focus on their duties. The enforcement of the punishment might remind Auditors of this expressive nature of the treatment.

	(1) Baseline, no misconduct	(2) Baseline, misconduct	(3) Treatment, no punishment	(4) Treatment, punishment
Dependent variable: Price awarded (1) or not (0).				
Deviation (standarized)	-1.40***	-1.61***	-1.31***	-2.01***
	(0.20)	(0.38)	(0.20)	(0.40)
Offered Transfer (standardized)	0.42***	0.39**	0.25**	0.26
	(0.10)	(0.17)	(0.12)	(0.19)
Constant	-1.65**	-1.31	-0.25	-2.59**
	(0.76)	(1.61)	(0.77)	(1.15)
Controls	Yes	Yes	Yes	Yes
Observations	679	236	762	183
Pseudo R^2	0.191	0.235	0.162	0.287

Logit regressions with clustered standard errors in parentheses
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table 4: Logit regressions of the award decision in the baseline and treatment separately by misconduct in previous round

The finding for expressive law might imply that this influence overlaps with that of crowding-out. We can test the simultaneity of these influences at the individual level. Some Auditors might experience a crowding-out effect and lose their intrinsic motivation. Others may be reminded to respect norms and focus their attention on their duties. To identify such different potential types of Auditors, we run regressions separately for each Auditor, based on the observations from the 16 rounds.[7] This yields individual estimates for each Auditor, indicating their intrinsic motivation and corrupt reciprocity. The relationship between these two estimates is shown in a scatterplot in figure 9. Each dot in the plot represents an Auditor with the coefficient of the deviation displayed on the y-axis and the coefficient of the transfer on the x-axis. Auditors who place a high weight on the transfer are located further to the right. To make the scatterplot easier to read, we multiply the coefficient of the deviation by (-1), so that highly motivated Auditors receive higher values than those with low intrinsic motivation, which makes them appear higher up.

We might expect that two types of Auditors dominate. There is the incorrupt and dutiful type and there is the corrupt and negligent type. A dominance of these two types would imply a negative relationship between the two coefficients. Some dots would be in the upper left corner, while others would be in the lower right corner. This cannot be confirmed for the baseline, shown in the left panel of figure 9. Auditors who give little weight to transfers, located

[7] Rather than using Logit regressions, as in table 2, we had to employ simple OLS because the 16 award decisions were in some cases fully explained by the explanatory variables.

to the left of the panel, are not characterised by higher intrinsic motivation. Auditors who focus on transfers, located to the right, do not show lower levels of intrinsic motivation. In the treatment, however, there appears to be a clear inverse relationship, as can be seen in the right panel of figure 9. Auditors who place little importance on transfers tend to show high levels of intrinsic motivation. Those Auditors for whom transfers are important are less intrinsically motivated.[8]

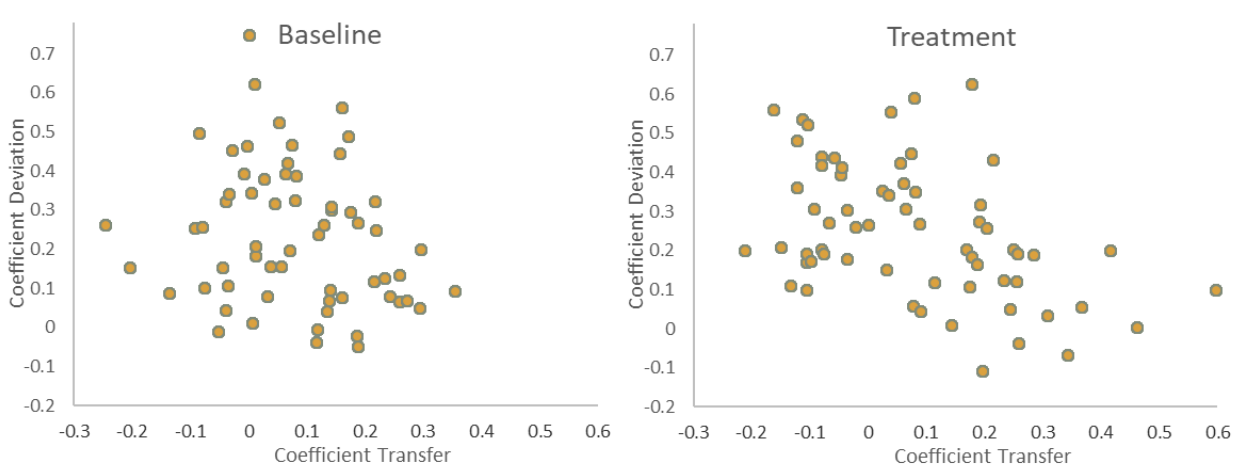


Figure 9: Scatterplots on coefficients for individual Auditors

5. Discussion and Conclusion

We ran a competition between two opposing views and identified the influence of corruption and intrinsic motivation on public sector performance and how it reacts to a standard form of punishment, i.e. punishing officials for accepting bribes. Our findings suggest a Solomonic balance between the two viewpoints. On the one hand, we show that intrinsic motivation, defined in rather broad terms, exerts a much higher influence on performance than corruption. This allows us to conclude that intrinsic motivation should receive more emphasis than methods to contain corruption. In a word, politicians should cultivate officials' intrinsic motivation and worry less about their corruptibility. Management tools that foster intrinsic motivation deserve more recognition than methods that reduce corruption.

On the other hand, we do not find a decrease in intrinsic motivation or an increase in type II errors in a treatment with punishment. Our findings thus suggest that punishing the acceptance of bribes can be implemented without harming motivation and without detrimental welfare effects in the form of type I or II errors. In essence, penalizing the act of accepting bribes is not inherently wrong, yet it falls short of yielding significant favorable effects.

We also find that there are two types of officials: Those whose intrinsic motivation is crowded out by punishment and those who react to the expressive content of a punishment with an increase in intrinsic motivation. In the public sector, in contrast to the private sector, officials often face punishment for taking gifts. The employment conditions in the real-world public sector are close to our treatment. This shows that it is particularly relevant for the public sector

[8] In Appendix D we report regression results on the scatterplots, revealing that the coefficients of transfers do significantly impact those of the deviation in the treatment but not in the baseline.

to recruit intrinsically motivated officials. These are likely to be a type who also values the expressive content of a punishment. One method to do so might be the recruitment of qualified personnel or internal promotion based on merit. It is likely that highly qualified officials are also the ones that exhibit high levels of intrinsic motivation. Selecting these would ensure that officials, on average, are less responsive to transfers and thus less corruptible. Arezki et al. (2012), for example, show that countries with highly qualified public officials perform better with regard to tax collection and financial sector regulation. In line with our results, they also find that these officials are characterized by lower levels of corruption. A focus on qualification may thus be a good way of achieving both, intrinsic motivation and a resistance to bribery.

Whether our experiment provides a fair comparison between intrinsic motivation and corruption is certainly a judgment call. There are arguments in favor of such a claim: We did not consciously manipulate the design to favor either side, but were open to both sides winning the contest. Our design is also in line with the established experimental literature on corruption and intrinsic motivation and well resembles real-world interactions between businesspeople and officials. Against these arguments, one could certainly vary parameters or design features and might obtain different results. Nevertheless, given that intrinsic motivation is more important than corruption by an order of magnitude, we are confident that such variations would face an uphill struggle in overturning our findings. Ultimately, however, we contend that this must remain an avenue for future research.

There are many ways to strengthen intrinsic motivation. Linking our findings to the broad interdisciplinary research from psychology, education, and management that discusses methods to enhance motivation would be beyond the scope of this article. There are some studies that address methods to increase honesty and compliance. For example, the "tone at the top" is regarded as an important determinant for compliance. Schwartz et al. (2005: 87) argue: "Boards sit at the top of the corporate hierarchy. Along with senior management, directors set by their words and deeds the ethical tone for the organizations". Kaptein and Wempe (1998: 862) even suggest that the tone at the top is the most important factor that impacts ethical behavior in organizations. d'Adda et al. (2017) investigate cheating in a die-rolling experiment and allow leaders to influence followers using incentives and statements. These statements, which may be described as a tone at the top, exert substantial influence. The statements of unethical leaders, as judged by their own behavior, induced more cheating among followers. Likewise, leaders who honestly played the die-rolling task submitted statements that enhanced honesty of their followers. Overall, the authors show that the tone at the top exerts an influence that is even more pronounced than that of extrinsic incentives. In a similar spirit, legal codes have increasingly focused on ethics codes within organizations. In the USA, the Sarbanes-Oxley Act requires companies to adopt a code of ethics for senior financial officers. The US Sentencing Guidelines state that an organization shall "romote an organizational culture that encourages ethical conduct and a commitment to compliance with the law."

The tone at the top might also promote intrinsic motivation in a broader sense. Communication and an appreciative relationship between managers and employees can play an important role here. An intrinsic motivation may be advanced if the broader goals of policies and duties are clearly communicated. The importance of acknowledgement and an appreciative leadership style is, for example, found in Erikkson and Villeval (2012). Subjects play a gift-exchange game, framed as a labor market. In one treatment, the employer can send between 0

and 5 symbolic rewards to the employee after observing the effort level. Effort significantly increases with this type of feedback. Kube et al. (2012) show how gifts by employers can promote the intrinsic motivation of employees. It is not only the monetary value of the gift that is relevant but also the time and effort the employer invests into the gift. Gifts that required time for preparation and signal effort strengthen the relationship and induce a type of reciprocity that benefits the employer. This might bring about suggestions for the public sector. Superiors should be given some opportunity to make gifts to the subordinates to express their appreciation and they should be expected to devote time and effort to such a task.

Our findings also have implications for risk management in the public sector – and the private sector alike. Corruption is commonly a risk to business because companies are often liable for the behavior of their employees and because corruption facilitates the leakage of resources to the detriment of the business. Our findings suggest that a lack of intrinsic motivation is the greatest risk. Surely, risk management should take corruption seriously. But risk management should not shift the focus in the wrong direction. Performance depends less on methods that reduce the risk of corruption and more on the joy and satisfaction that employees experience in their work environment. It is more important to select those who are intrinsically motivated, rather than those who abstain from risky interactions, even if they are posted to sectors or countries where they face high levels of corruption.

The Solomic balance in our findings can also be interpreted from an organizational perspective. It is standard to separate management from control. Such a separation of tasks need not bring about conflicts. For example, the efforts by law enforcers, internal and external reviewers or compliance may focus on detection and punishment. Given that we do not find evidence on crowding out at the aggregate level, such efforts do not necessarily impair those of managers who focus on motivating their staff.

Our experiment has been conducted in Germany, a country perceived to be less affected by corruption and where intrinsic motivation may be strong. It remains to be seen whether our findings can be replicated in environments with rampant corruption. In such environments, the monetary incentives to engage in corruption are large and people might expect a bribe in exchange for reciprocity. This is also emphasized by Bosio et al. (2022), who investigate public procurement and find that in countries with low public sector capacity, which tend to be poor countries with higher levels of corruption, stricter laws improve outcomes. On the other hand, in rich countries with high public sector capacity, the role of ethics is strong, such that granting discretion to officials improves procurement outcomes. At the moment, thus, our findings may be externally valid only for environments with low or moderate levels of corruption. How the methods suggested here perform in countries and sectors that are prone to corruption should be an avenue for future research.

Acknowledgement

We thank the German Research Foundation (DFG; project number 448261984) for funding the project. We are grateful for critical comments made by Stefan Bauernschuster, Stephan Geschwind, Marcus Giamattei, Michael Grimm, Susanna Grundmann, Sebastian Krautheim, Rosemarie Nagel, Alexander Wagner, as well as by the audiences of the 1st Passau ECON

workshop, the 2022 meeting of the German Association for Experimental Economic Research (GfeW), and the 2023 CCC (CBESS-CEDEX-CREED) meeting.

References

Abbink, K. (2004) Staff rotation as an anti-corruption policy: an experimental study. *European Journal of Political Economy* 20: 887–906

Abbink, K. and H. Hennig-Schmidt (2006) Neutral versus Loaded Instructions in a Bribery Experiment. *Experimental Economics* 9: 103-121.

Abbink, K., B. Irlenbusch, and E. Renner, E. (2002) An Experimental Bribery Game. *Journal of Law, Economics and Organization* 18(2): 428-54.

Akerlof, G. and W. Dickens (1982) The Economic Consequences of Cognitive Dissonance. *American Economic Review* 72(3): 307-319

Anechiarico, F. and J. B. Jacobs. 1996. The Pursuit of Absolute Integrity. How Corruption Control Makes Government Ineffective. Chicago: University of Chicago Press.

Arezki, R., H. Lui, M.G. Quintyn, and F.G. Toscani (2012) Education Attainment in Public Administration Around the World: Evidence from a New Dataset. IMF Working Paper 2012/231

Arlen, J. and L.A.Kornhauser (2022) Battle for Our Souls: A Psychological Justification for Corporate and Individual Liability for Organizational Misconduct. Available at SSRN: https://ssrn.com/abstract=4152960

Bandiera, O., A. Prat, and T. Valletti (2009) Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review* 99(4):1278–308.

Bandiera, O., M.C. Best, A.Q. Khan, and A. Prat (2021) The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats. *The Quarterly Journal of Economics* 136(4): 2195–2242.

Bartling, B., E. Fehr, and K. M. Schmidt (2012) Screening, Competition, and Job Design: Economic Origins of Good Jobs. American Economic Review 102 (2): 834-64.

Bateson, M., D. Nettle, and G. Roberts (2006) Cues of being watched enhance cooperation in a real-world setting. *Biological Letters* 2: 412–414.

Becker, G. S. and G.J. Stigler (1974) Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Legal Studies* 3: 1-19.

Benabou, R., and J. Tirole (2003) Intrinsic and extrinsic motivation. *Review of Economic Studies*, 70(3), 489–520. doi:10.1111/roes.2003.70.issue-3

Benabou, R., and J. Tirole (2006) Incentives and Prosocial Behavior. *American Economic Review* 96(5): 1652-1678.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2015) "Does Management Matter in Schools? *Economic Journal* 125: 647–674.

Bolton, G. E. and R. Zwick (1995) Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behavior* 10: 95–121.

Bosio, E., S. Djankov, E. Glaeser, and A. Shleifer (2022) Public Procurement in Law and Practice. *American Economic Review* 112 (4): 1091-1117.

Bowles, S. (2008) Policies Designed for Self-Interested Citizens May Undermine 'the Moral Sentiments': Evidence from Economic Experiments. Science (5883):1605-9 .

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

Carr, M. D., and P. Mellizo (2013). The relative effect of voice, autonomy, and the wage on satisfaction with work. *The International Journal of Human Resource Management*, *24*(6), 1186-1201.

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

Cohn, A., M. Maréchal, D. Tannenbaum and C. Zünd (2019), Civic honesty around the globe. *Science* 365: 70-73.

Coviello, D., A. Guglielmo, and G. Spagnolo (2018) The Effect of Discretion on Procurement Performance. *Management Science* 64(2): 715–38.

d'Adda, G., D. Darai, N. Pavanini and R. A. Weber (2017) "Do leaders affect ethical conduct?" Journal of the European Economic Association 15(6):1177–1213.

Deci, E. L., R. Koestner, and R.M. Ryan (1999) A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin 125:627–668.

Dickinson, D. L., and M.C. Villeval (2008) Does monitoring decrease work effort? The complementarity between agency and crowding-out theories. *Games and Economic Behavior* 63: 56-76.

Dufwenberg, M. and G. Kirchsteiger (2004) A Theory of Sequential Reciprocity. *Games and Economic Behavior* 47 (2): 268–298.

Eriksson, T. and M.C. Villeval (2012) "Respect and relational contracts," Journal of Economic Behavior & Organization, Elsevier, 81(1): 286-298.

Falk, A. and M. Kosfeld (2006) The Hidden Costs of Control. *American Economic Review* 96 (5):1611-1630.

Fehr, E. and A. Falk (2002), Psychological foundations of incentives. European Economic Review 46 (2002) 687 – 724.

Fehr, E. and B. Rockenbach (2003) Detrimental effects of sanctions on human altruism. *Nature* 422: 137-140.

Fischbacher, U. and F. Föllmi-Heusi (2013) Lies in disguise – an experimental study on cheating. *Journal of the European Economic Association* 11(3): 525-547.

Frank, B. and J. Graf Lambsdorff (2010) Bribing versus Gift-Giving – an Experiment. *Journal of Economic Psychology* 31(3): 347-357.

Gächter, S. and J. Schulz (2016) Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531: 496-499.

Galeotti, F., V. Maggian, and M.C. Villeval (2021) Fraud Deterrence Institutions Reduce Intrinsic Honesty. *The Economic Journal* 131(638): 2508–2528.

Gneezy, U. and A. Rustichini (2000) Pay enough or don't pay at all. *Quarterly Journal of Economics* 115(3): 791-810.

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

Gneezy, U., S. Saccardo, and R. van Veldhuizen (2019) Bribery: Behavioral Drivers of Distorted Decisions. *Journal of the European Economic Association* 17(3): 917–946.

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

Grundmann, S. and J. Graf Lambsdorff (2017) How income and tax rates provoke cheating – An experimental investigation of tax morale. *Journal of Economic Psychology* 63: 27-42.

Güth, W. and M.G. Kocher (2013) More than Thirty Years of Ultimatum Bargaining Experiments: Motives, Variations, and a Survey of the Recent Literature. *CESifo Working Paper Series* 4380.

Güth, W., R. Schmittberger, and B. Schwarze (1982) An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization* 3, 367-388.

Gutman, J. (2014) Is there Room for Discretion? Reforming Public Procurement in a Compliance-Oriented World. Global Economy & Development Working Paper 74. Washington D.C.: Brookings Institution.

Hallsworth, M., J. A. List, R. D. Metcalfe and I. Vlaev (2017) The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics* 148, 14-31

Kaptein, M., and J. Wempe (1998) Twelve Gordian Knots When Developing an Organizational Code of Ethics. *Journal of Business Ethics* 17:853-869.

Kelman, S. (2003) Remaking Federal Procurement. *The John F. Kennedy School of Government Working Paper* 3 (Cambridge, MA).

Khadjavi, M. (2014) On the Interaction of Deterrence and Emotions. *The Journal of Law, Economics, and Organization* 31(2): 287–319.

Kube, S., M. A. Maréchal, and C. Puppe (2012) The Currency of Reciprocity: Gift Exchange in the Workplace. *American Economic Review* 102(4):1644-62.

Lambsdorff, J. Graf (2007) The Institutional Economics of Corruption and Reform: Theory, Evidence, and Policy. Cambridge University Press.

Lambsdorff, J. Graf and B. Frank (2011) Corrupt Reciprocity – Experimental Evidence on a Men's Game. *International Review of Law and Economics* 31 (2): 116-125.

Lane, T., D. Nosenzo and S. Sonderegger (2023) Law and Norms: Empirical Evidence. *American Economic Review* 113(5): 1255-1293.

Nosenzo, D., E. Xiao, and N. Xue (2022) Norm-Signalling Punishment. Economics Working Papers 2022-07, Aarhus University.

Patall, E. A., H. Cooper, and J.C. Robinson (2008) The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin* 134(2): 270–300.

Polinsky, A.M. and S. Shavell (2000) The Economic Theory of Public Enforcement of Law. *Journal of Economic Literature* 38(1): 45-76.

Polinsky, A.M. and S. Shavell (2001) Corruption and optimal law enforcement. *Journal of Public Economics* 81: 1–24.

Pruckner, G. and R. Sausgruber (2013) Honesty on the streets: a field study on newspaper purchasing. *Journal of the European Economic Association* 11(3):661 – 679.

Rabin, M. (1993) Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83 (5): 1281–1302.

Rasul, I., and D. Rogger (2016) Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *Economic Journal*, 128: 413–446.

Rasul, I., D. Rogger, and M.J. Williams (2021) Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service. *Journal of Public Administration Research and Theory* 31: 259–277.

Rivas, F. (2013) An Experiment on Corruption and Gender. Bulletin of Economic Research 65(1): 10-42.

Rose-Ackerman, S. and R. Truex (2012) Corruption and Policy Reform. Yale Law & Economics Research Paper 444

Schildberg-Hörisch, H. and C. Strassmair (2012) An Experimental Test of the Deterrence Hypothesis. *The Journal of Law, Economics, and Organization* 28(3): 447–459

Schulze, G., and B. Frank (2003): Deterrence versus Intrinsic Motivation - Experimental Evidence on the Determinants of Corruptibility, *Economics of Governance* 4, 143-160.

Schwartz, M. S., T.W. Dunfee, and M.J. Kline (2005) Tone at the top: An ethics code for directors? *Journal of Business Ethics*. 58: 79–100

Serra, D. and L. Wantchekon (2012) New Advances in Experimental Research on Corruption. Research in Experimental Economics Vol. 15. Bingley: Emerald Books.

Shalvi, S., J. Dana, M.J.J.Handgraaf, and K.W.De Dreu (2011) Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* 115: 181–190.

Spagnolo, G. (2012) Reputation, Competition and Entry in Procurement. *International Journal of Industrial Organization* 30 (3): 291–96.

Sunstein, C. R. (1996). On the expressive function of law. *University of Pennsylvania law review* 144(5):2021-2053.

Tanner, C., S. Linder, and M. Sohn (2022) Does moral commitment predict resistance to corruption? experimental evidence from a bribery game. PLoS ONE 17(1): e0262201.

Thomas, K. W. (2009) Intrinsic Motivation at Work: What Really Drives Employee Engagement. 2nd ed. Oakland, USA: Berrett-Koehler Publishers.

United Nations (2004), United Nations Convention Against Corruption. Vienna International Centre: Vienna, Austria.

A. Experimental instructions

The Appendix reproduces all instructions for the baseline. Treatment instructions differed from the baseline by substitution of one sentence as highlighted on page 35. General instructions (A.1) regarding the purpose of the experiment, its expected length, the conversion of the experimental currency unit talers into euros, the rules of behavior, and the (standard) payment and blindness procedures were given orally. The core instructions of the experiment (A.2) were shown on the screen of the computer. Comprehension questions (A.3) had to be solved correctly to proceed. After the experiment concluded, subjects were asked to complete a post-experimental questionnaire (A.4). Original instructions were given in German.

A.1 General instructions

Welcome to this Experiment! Before the experiment begins, you will receive some general information.

The aim of the experiment is to obtain insights into human behavior. The participants in the experiment are all here in the room and are taking part in the same experiment.

At the end of the experiment, you will in any case receive 3.50€ for showing up on time. In addition, you can earn money in the experiment, which will be expressed in Talers. How much you can earn exactly depends on your decisions as well as the decisions of the other participants. At the end of the experiment, your earned Talers will be converted into Euro. 1 Taler equals 20 euro cents.

All participants are anonymous and cannot communicate with each other. The decisions you make and the information you provide will not be linked to you as a person.

The payment as well will be anonymous: no other participant will see how much you receive, and the experimenter will not know this either. The person that pays you the money at the end does not know about the experiment and cannot draw any conclusions about your behavior from the amount you have earned.

The procedures of the experiment will be explained to you on the screens to be displayed in a moment. Please read all the information carefully. After reading the instructions, you will be asked to answer comprehension questions. If you answer them correctly on the first attempt, you can earn additional Talers.

Please note that you may experience waiting times during the experiment. Please remain quietly seated at your desk throughout the experiment and refrain from any conversations. Violations of these rules may result in exclusion from further participation in the experiment.

Since the experiment will be conducted over several days, we hereby inform you that any communication about the content of the experiment (e.g., via Jodel) is prohibited. Such communication violates the requirements of the laboratory and may result in experiments being no longer conducted in the future.

If you have any questions, please raise your hand. Someone will then come to you.

You can now start the experiment.

A.2 Core instructions

Instructions

Welcome to this experiment!

The experiment consists of 16 rounds. At the beginning of the experiment, each participant will be randomly assigned the role of the **Estimator** or the **Auditor**. The assigned role will remain the same throughout the 16 rounds of this experiment. At the beginning of every round, one Estimator and one Auditor will be randomly and anonymously matched to interact with each other. Neither the Estimator nor the Auditor will interact with the same person in consecutive rounds. In every round, the Estimator and the Auditor can obtain positive or negative income, which is expressed in **Talers**. Every round consists of three stages, which are explained on the following pages.

Next

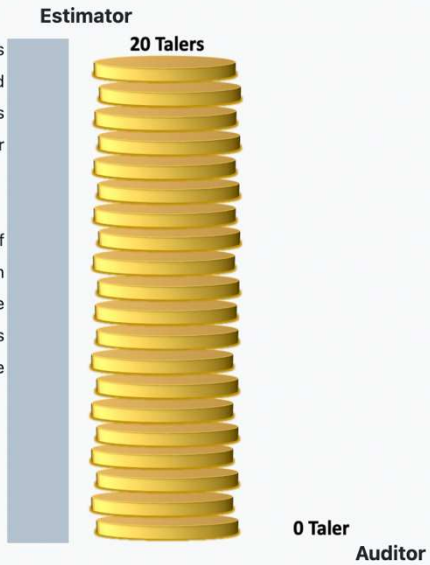
Instructions

Stage 1: Transfer

In every round, the Estimator starts with an endowment of 20 Talers and **proposes** how much of this endowment to transfer to the Auditor by positioning a slider.

Please put yourself in the position of the Estimator and click anywhere in the grey area of the graph on the right. Move the slider that appears and confirm your proposal using the following button:

Confirm Proposal



Back

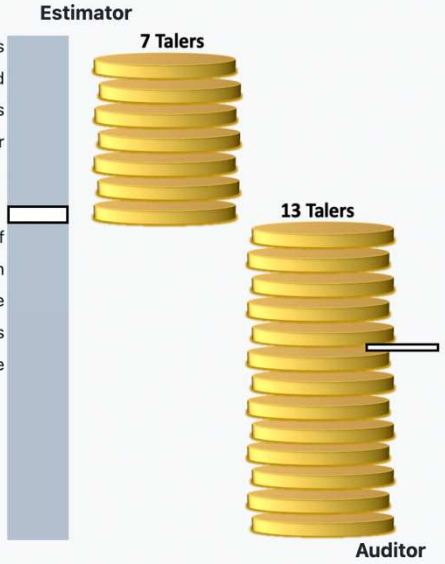
Instructions

Stage 1: Transfer

In every round, the Estimator starts with an endowment of 20 Talers and **proposes** how much of this endowment to transfer to the Auditor by positioning a slider.

Please put yourself in the position of the Estimator and click anywhere in the grey area of the graph on the right. Move the slider that appears and confirm your proposal using the following button:

Confirm Proposal



If the Auditor **accepts** the proposal, it is implemented. If the Auditor **rejects** the proposal, the Estimator keeps 12 Talers and the Auditor receives 8 Talers (see slim bar).

Please put yourself in the position of the Auditor and for practice alternatingly press each of the following buttons:

Accept Reject

Back

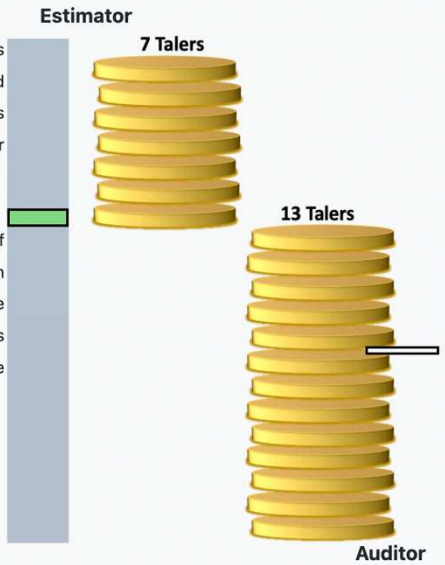
Instructions

Stage 1: Transfer

In every round, the Estimator starts with an endowment of 20 Talers and **proposes** how much of this endowment to transfer to the Auditor by positioning a slider.

Please put yourself in the position of the Estimator and click anywhere in the grey area of the graph on the right. Move the slider that appears and confirm your proposal using the following button:

Confirm Proposal



If the Auditor **accepts** the proposal, it is implemented. If the Auditor **rejects** the proposal, the Estimator keeps 12 Talers and the Auditor receives 8 Talers (see slim bar).

Please put yourself in the position of the Auditor and for practice alternatingly press each of the following buttons:

Accept Reject

Back

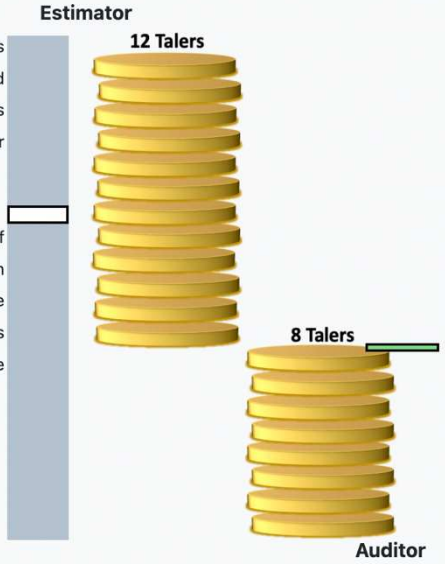
Instructions

Stage 1: Transfer

In every round, the Estimator starts with an endowment of 20 Talers and **proposes** how much of this endowment to transfer to the Auditor by positioning a slider.

Please put yourself in the position of the Estimator and click anywhere in the grey area of the graph on the right. Move the slider that appears and confirm your proposal using the following button:

Confirm Proposal



If the Auditor **accepts** the proposal, it is implemented. If the Auditor **rejects** the proposal, the Estimator keeps 12 Talers and the Auditor receives 8 Talers (see slim bar).

Please put yourself in the position of the Auditor and for practice alternatingly press each of the following buttons:

Accept Reject

Back Next

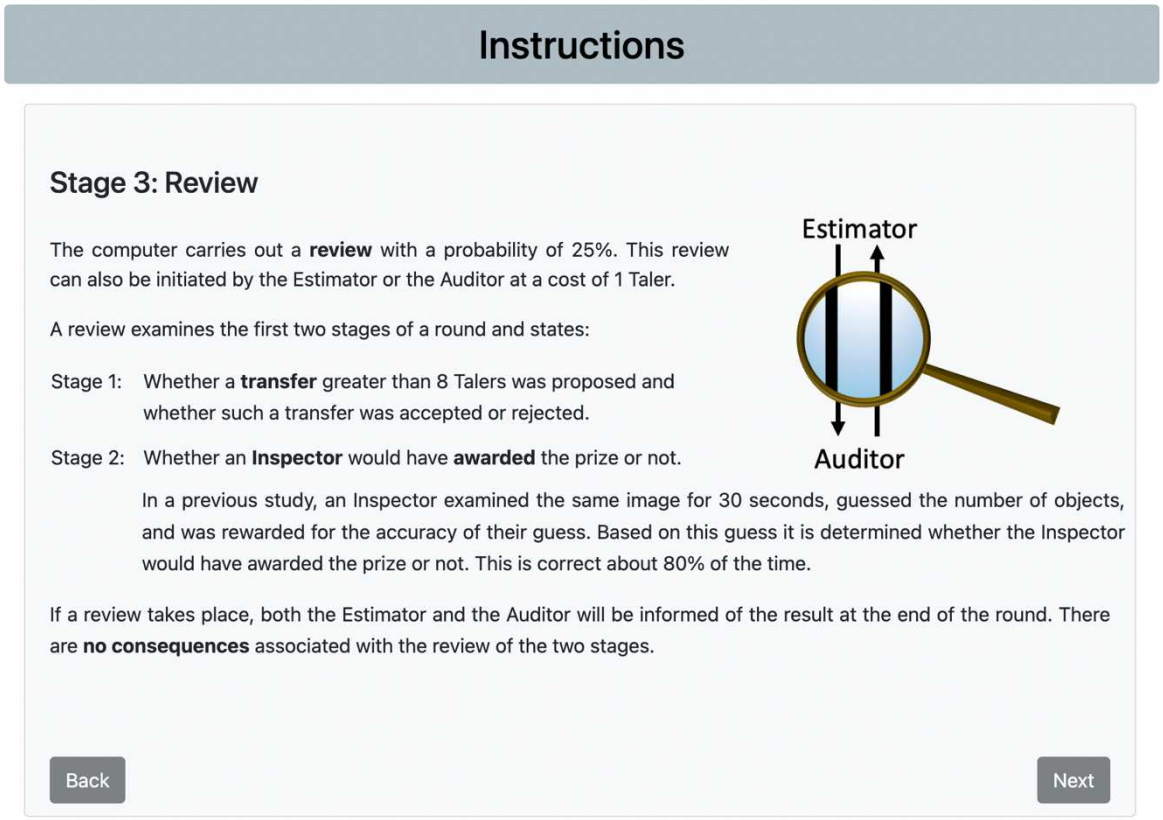
Instructions

Stage 2: Guess and award prize

Next, the Estimator will be asked to **guess** the number of objects in an image, shown for 6 seconds. An example image with 76 objects is provided on the right.

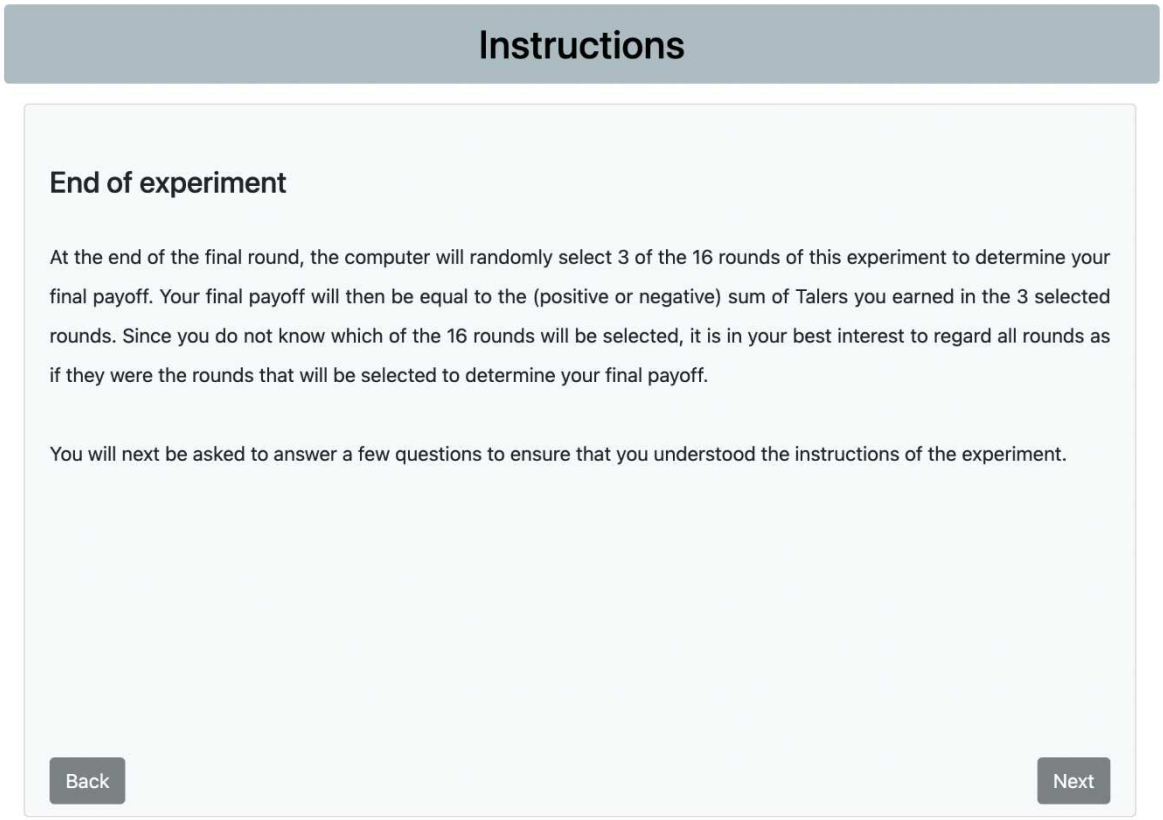
The Auditor sees the same image for 30 seconds and decides whether the Estimator's guess is good enough to deserve a 20 Talers prize. The task is to **award** the prize if the Auditor believes that the Estimator's guess differs by 5 or less from the correct number of objects. The task is **not to award** the prize if the Auditor believes that the Estimator's guess differs by more than 5 from the correct number of objects.

Back Next



Note: Our treatment instructions replaced the last sentence in the above screen (i.e. 'There are **no consequences** associated with the review of the two stages.') as follows:

Treatment: 'The Auditor will have **8 Talers deducted from their income** if the review of stage 1 detects that the Auditor accepted a transfer greater 8. There are no consequences associated with the review of stage 2.'



A.3 Comprehension questions

Subjects faced 7 comprehension questions which had to be passed before they were able to proceed. They received 1 taler for each comprehension question they answered correctly in the first attempt. While answering the comprehension questions, subjects were able to return to the instructions using back and forth buttons. Filled squares indicate solutions for the baseline. Treatment questionnaires used the same combination of questions and answers as the baseline and merely differed in the correctness of responses to questions 5 and 6 as indicated below.

Question 1: Think of any possible transfer proposal. Suppose the Auditor decides to reject the transfer proposal. What are the consequences of this?

☐ The Estimator keeps 0 Taler, the Auditor receives 0 Taler.

■ The Estimator keeps 12 Taler, the Auditor receives 8 Taler.

☐ The Estimator keeps 20 Taler, the Auditor receives 0 Taler.

☐ The Estimator keeps 8 Taler, the Auditor receives 12 Taler.

Question 2: Suppose the Auditor believes to have seen 46 objects. Under what circumstances should the Auditor award the prize to the Estimator?

☐ The Auditor should award the prize if a transfer greater 8 was proposed.

☐ The Auditor should award the prize if the Estimator guessed 54 objects.

■ The Auditor should award the prize if the Estimator guessed 48 objects.

☐ The Auditor should award the prize if the Estimator guessed 40 objects.

Question 3: Suppose neither the Estimator nor the Auditor decide to initiate a review. What is the consequence of this?

☐ A review will not be carried out.

■ There is a probability of 25% that the computer will carry out a review.

☐ A review will always be carried out.

☐ There is a probability of 75% that the computer will carry out a review.

Question 4: Suppose a transfer of 10 is proposed and accepted. The Auditor decides <u>not</u> to award the prize. What is the consequence if <u>no</u> review is carried out?

■ The Estimator earns 10 Taler, the Auditor earns 10 Taler.

☐ The Estimator earns 10 Taler, the Auditor earns 2 Taler.

☐ The Estimator earns 10 Taler, the Auditor earns 30 Taler.

☐ The Estimator earns 2 Taler, the Auditor earns 10 Taler.

Question 5: Suppose a transfer of 18 is proposed and accepted. The Auditor decides to award the prize. What is the consequence if a review by the computer in addition detects that the Inspector would have awarded the prize?
[**Note:** Treatment differences marked in square brackets.]

■ The Estimator earns 22 Taler, the Auditor earns 18 Taler.

☐ The Estimator earns 22 Taler, the Auditor earns 10 Taler. [Correct in the treatment]

☐ The Estimator earns 14 Taler, the Auditor earns 10 Taler.

☐ The Estimator earns 14 Taler, the Auditor earns 18 Taler.

Question 6: Suppose a transfer of 15 is proposed and accepted. The Auditor decides to award the prize. What is the consequence if a review by the computer in addition detects that the Inspector would not have awarded the prize?
[**Note:** Treatment differences marked in square brackets.]

☐ The Estimator earns 17 Taler, the Auditor earns 15 Taler.

☐ The Estimator earns 25 Taler, the Auditor earns 7 Taler. [Correct in the treatment]

■ The Estimator earns 25 Taler, the Auditor earns 15 Taler.

☐ The Estimator earns 5 Taler, the Auditor earns 15 Taler.

Question 7: In every round of the experiment you can earn positive or negative income expressed in Talers. The final payoff is the sum of Talers you earned in …

☐ each of the 16 rounds. Rounds with negative income can affect your final payoff.

■ 3 randomly selected rounds. Rounds with negative income can affect your final payoff.

☐ 3 randomly selected rounds with positive income. Rounds with negative income cannot affect your final payoff.

☐ all rounds with positive income. Rounds with negative income cannot affect your final payoff.

A.4 Post-experimental questionnaire

Upon completion of the experiment, but before subjects were informed of the rounds selected to determine their payoffs, they were asked to complete the following role-specific questionnaire. Each questionnaire entailed two incentivised belief elicitation questions which were asked on separate screens to avoid contamination.

Estimator questionnaire:

Question 1: What criteria did you use in general to determine your transfer proposals and what did you hope to gain from your decisions? Type here.

Question 2: (Bonus: If your assessment is correct, you will receive 1 Taler) **Compared to all other Estimators, do you think your guesses over the 16 rounds were *less often*, *equally often* or *more often* good (i.e., worthy of a prize)?** (A difference smaller than 1 compared to the average Estimator counts as equally often.)

- less often
- equally often
- more often

Question 3: (Bonus: If your assessment is correct, you will receive 2 Talers) **You were awarded the prize x times over the 16 rounds. In your opinion, how often did you objectively deserve the prize (with an allowed deviation of not more than 5)?**

- 0 times
- …
- 16 times

Question 4: What made you decide to request or not to request a review? What thoughts were going through your mind? Type here.

Question 5: How old are you?

Question 6: Which gender do you identify with?

- male
- female
- diverse
- prefer not to say

Question 7: Which faculty do you feel most closely connected to?

Question 8: Have you experienced any technical problems or do you have any other comments? Type here.

Auditor questionnaire:

Question 1: What criteria did you use in general to decide whether to accept or reject transfer proposals? Type here.

Question 2: What criteria did you use in general to decide whether to award or not award the 20 Talers prize? Type here.

Question 3: (Bonus: If your assessment is correct, you will receive 1 Taler) **Compared to all other Auditors, do you think that over the 16 rounds you awarded the prize *less often*, *equally often* or *more often*?** (A difference smaller than 1 compared to the average Auditor counts as equally often.)

- less often
- equally often
- more often

Question 4: (Bonus: If your assessment is correct, you will receive 2 Talers) **You awarded the prize x times over the 16 rounds. In your opinion, how often would you objectively have had to award the prize (with an allowed deviation of no more than 5)?**

- 0 times
- …
- 16 times

Question 5: What made you decide to request or not to request a review? What thoughts were going through your mind? Type here.

Question 6: How old are you?

Question 7: Which gender do you identify with?

- male
- female
- diverse
- prefer not to say

Question 7: Which faculty do you feel most closely connected to?

Question 8: Have you experienced any technical problems or do you have any other comments? Type here.

B Model Specifications

Auditors can err in one of two directions: by awarding a prize to an undeserving Estimator, they commit a type I error; by not awarding a prize to a deserving Estimator, they commit a type II error. Figure B.1 depicts the expected relationship between an Auditor's level of tolerance and the probabilities of committing either of the errors or none. The shown data is based on simulations of random encounters between Estimators and unbiased Auditors from the aforementioned pilot study from which we also sampled Inspectors.

The results show that zero tolerance ($\bar{\tau} = 0$) implies both types of errors to be roughly balanced, such that neither error dominates. As the tolerance increases (in the experiment, this is induced by an increased transfer), the probability of withholding the prize is expected to decrease. Both the probability of correctly withholding the prize as well as that of a type II error, i.e., an incorrect withholding, decrease as well. The probability of awarding the prize, both correctly and incorrectly, is expected to increase with the transfer.

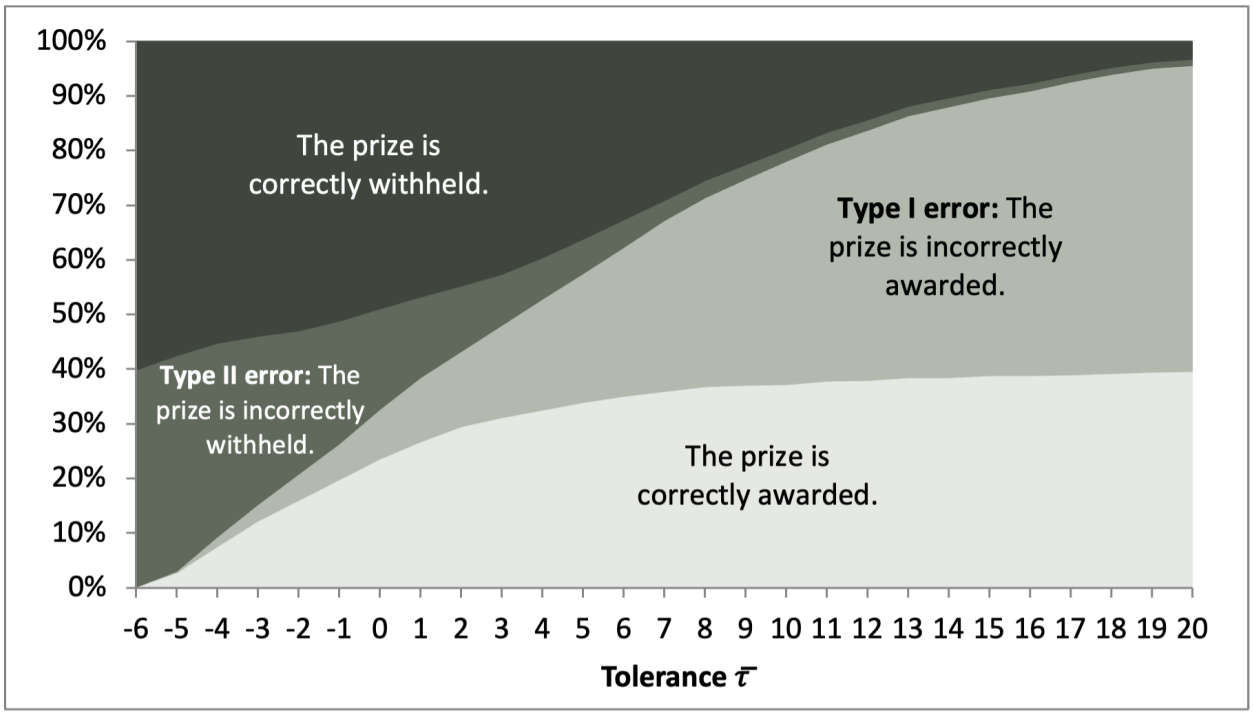


Figure B.1: Probability of objectively correct decision and errors of type I and type II

Estimator predictions

The Estimator will weigh the costs of a transfer against the returns from increased tolerance. For this, he will form expectations regarding the distributions of guesses formed in the task by Estimators and Auditors. For the probability π of an award of the prize, the function $\pi = \pi(\bar{\tau})$ holds with $\pi' > 0$ and $\pi'' < 0$, just as in figure B.1. The profit of the Estimator thus follows as:

$$P = 20 - T + 20 \cdot \pi \left(\frac{\lambda}{c} 20(T - 10)\right).$$

An optimum requires $P' = 0 \Leftrightarrow -1 + 20 \cdot \pi' \cdot \frac{\lambda}{c} 20 \Leftrightarrow 400 \frac{\lambda}{c} \pi' = 1$. Thus, for exemplary values $\frac{\lambda}{c} = 0.05$, it holds that $\pi' = \frac{1}{20}$. In the optimum, an increase of the transfer by 1, which brings about a marginal increase in tolerance also by 1, should increase the probability of the prize being awarded by $\frac{1}{20} = 5\%$ and, thus, of expected return also by 1 taler.

In figure B.2, the isocost curve has a slope of 1, in line with $\frac{\lambda}{c} 20$=1. An Auditor with zero tolerance will assess the Estimator's guess in an unbiased fashion and award the prize with probability $\pi(0) = \frac{1}{3}$ (according to the results of the simulation above). The expected return of receiving the prize is thus $20\frac{1}{3} = 6.66$ talers which corresponds to one point on the curve for the expected return. As the transfer increases, so does $\bar{\tau}$ and therefore π. The closer π approaches 1, the smaller the impact of additional tolerance becomes. This explains the concave shape of the expected returns. The isocost curve, depicted as a straight line in the figure, has a tangential point with the concave curve of returns. This tangential point marks the optimum at $\bar{\tau} = 3$ and $T = 10$.

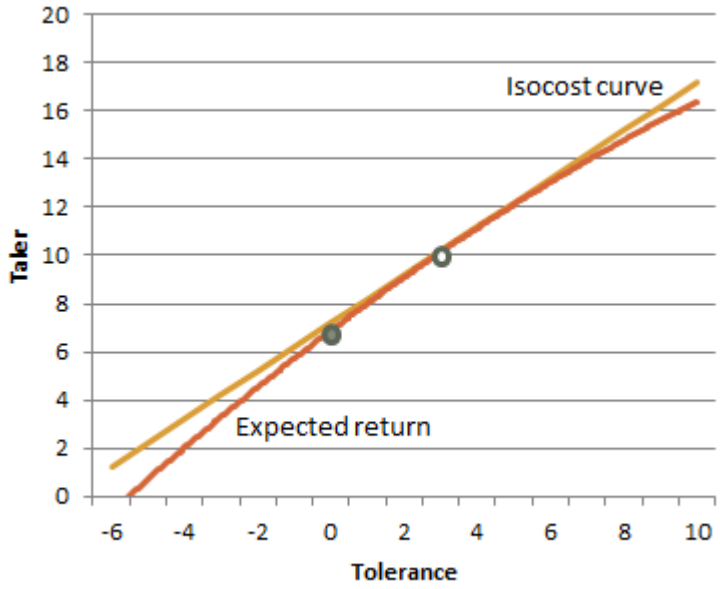


Figure B.2: Expected return of prize award and costs of tolerance.

C Design Details

C.1 Picture selection

The selection of pictures used in the experiment followed careful procedures. We were interested in pictures that were engaging and satisfied pre-specified characteristics which we discuss in more detail below. Starting off with 57 potential pictures, we first, in an informal way, let a total of 10 friends and relatives rate the originality and attractiveness of each picture on a scale from 1 (= original/attractive, I enjoy the look of the picture) to 5 (= not very original/not very attractive, I don't enjoy the look of the picture). After having reduced our

sample of pictures to the 40 most appealing ones, we conducted another informal online study which let a different total of 50 friends and relatives estimate the number of objects in each picture under a time-constraint of 6 seconds. This study was meant to give us a feel for the difficulty of the task, allowing us to replace/fine-tune some of the pictures, and to detect potential technical challenges (e.g. regarding the visualization and the display-length of pictures).

Our informal tests were followed by three incentivized pilots carried out with a UK sample totaling 160 subjects, recruited via the platform Prolific (www.prolific.co). The first two pilots let subjects face each of the remaining 40 pictures in a random order and asked them to estimate the number of objects in each picture – these two pilots merely differed in the display-length of each picture which was either 6 seconds or 30 seconds. Subjects were rewarded for the accuracy of their guesses for 3 randomly selected pictures. We used the data generated in these pilots to boil the number of pictures down to the 16 that we used in the experiment. We send out the complete list of pictures to researchers who seek to replicate our findings and keep them confidential from others in order to preserve their unbiased use in future experiments.

By design, we were looking for pictures that could reasonably be expected to generate a balance between Type I and Type II errors in the Auditor's award decision, under the assumption of 0 tolerance (i.e. $\tau = 0$) – we judged such circumstances to be ideal for our aim of creating room for discretion. To identify appropriate pictures, we first simulated random encounters between guesses submitted by Estimators (in the 6 seconds pilot) and Auditors (in the 30 seconds pilot) and determined whether $|s_{it} - p_{it}| \leq 5 + \tau$. A comparison with the objectively correct value then allowed us to determine which error (if any) occurred. Figure B.1 presents simulation results for a selected picture – the remaining 15 selections closely resembled the shown patterns.

Finally, it is possible that an estimate may serve as an anchor and thereby induce an Auditor to some intrinsic tolerance. As an example, after having received a high estimate, the Auditor overestimates the test value. To determine the extent and significance of such an anchor effect, in our third and final pilot we made a slight adjustment to the 30 seconds pilot by letting Auditors receive random guesses generated in the 6 seconds pilot. This pilot has been used to establish a first-best case, a case where Auditors are extrinsically motivated to make good guesses. The results are shown in the main text in figure 4, referred to as "extrinsic motivation". Anchor effects are relevant for this first-best case as well as for the baseline and all subsequent treatments. None of the differences observed in this study can thus be due to this anchor effect.

D Further Regression Analysis

Regressions in table 2 use the offered transfer as an explanatory variable. One might contend that the Auditor is not so much influenced by the offer but by the transfer ultimately received. If a transfer is rejected, it might not influence the award decision. While there is some truth to this argument, it would raise concerns regarding endogeneity. An Auditor may simultaneously decide on the response to the transfer and the decision to award. The realized transfer is then not fully exogenous to a regression on the award decision. Data on the offered transfer, to the contrary, do not raise such concerns as they are determined by the Estimator and therefore exogenous. This motivated the choice of the variable in table 2. We nonetheless report regressions here based on the realized transfer. While admitting problems of endogeneity, the results remain insightful because they confirm the robustness of the approach. As can be seen in table D.1, the coefficients are close to those reported in tables 2 and 3.

	(1) Baseline	(2) Treatment
Dependent variable: Price awarded (1) or not (0)		
Deviation (standarized)	-1.46***	-1.46***
	(0.2)	(0.2)
Realized Transfer (standardized)	0.46***	0.40***
	(0.09)	(0.1)
Constant	-2.05***	-0.75
	(0.7)	(0.7)
Controls	Yes	Yes
Observations	976	1008
Pseudo R^2	0.199	0.184

Logit regressions with clustered standard errors in parentheses
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table D.1: Logit regressions of the award decision with realized rather than offered transfers

When registering our experiment (aspredicted.org, #80975), we committed to leaving out subjects who performed poorly on the comprehension questions and had less than 4 correct answers from a total of 7 questions. In tables 2 and 3 we decided nonetheless to include all subjects. This was motivated by verbal feedback from many subjects who admitted that they did not fully understand the instructions in the beginning but got to understand the game after the first two rounds. Given our commitment at the time of registration, we here report the regressions without the Auditors who failed the comprehension questions. As can be seen in table D.2, the results remain largely unaffected.

	(1) Baseline	(2) Treatment
Dependent variable: Price awarded (1) or not (0)		
Deviation (standarized)	-1.49***	-1.44***
	(0.2)	(0.2)
Offered Transfer (standardized)	0.44***	0.55***
	(0.09)	(0.1)
Constant	-2.10***	-0.79
	(0.7)	(0.7)
Controls	Yes	Yes
Observations	848	864
Pseudo R^2	0.203	0.187

Logit regressions with clustered standard errors in parentheses
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table D.2: Logit regressions of the award decision in the baseline without Auditors who failed the comprehension questions

We can confirm the relationship displayed in figure 8 by help of regressions, as shown in table D.3, where we regress the coefficient we obtained for the deviation on the coefficient we obtained for the transfer across all Auditors. Each Auditor represents one observation. The regressions confirm the relationship we found in figure 8: While there is no significant relationship between the effect of the deviation and the transfer on the award probability in the baseline, we see an inverse relationship in the treatment.

	(1) Baseline	(2) Treatment
Dependent variable: coefficient of deviation of each Auditor		
Offered Transfer (standardized)	-0.27	-0.40***
	(0.2)	(0.1)
Constant	0.20	0.45***
	(0.2)	(0.1)
Controls	Yes	Yes
Observations	61	63
R^2	0.082	0.247

OLS regression with one coefficient per subject
Controls: age, gender, round dummies, location
* p<0.10, ** p<0.05, *** p<0.01

Table D.3: OLS regressions of the coefficient of the deviation on the coefficient of the transfer, separately for baseline and treatment

E Power Test

We ran a power-test with conventional values for β (0.2) and α (0.05) to assess the necessary sample size. Crucial variables of interest are binary variables such as the Auditor's decision to reject a transfer, to award the prize or to commit a type I error. The treatment should at least affect the probability of awarding the prize by 10 percentage points, for example increasing it from 30 to 40 percent. Given a binomial distribution, the standard error would be $\sqrt{0.3 * 0.7} \cong 0.46$. We planned (and implemented) identical sample sizes for treatment and baseline and consequently a sampling ratio of 1. A power test sets the minimum to 332 observations, which would translate to $\frac{142}{16} = 9$ Auditors under the unrealistic assumption that their choices are completely independent across rounds.

Dropping this unrealistic assumption, we follow List et al. (2011) and Moffat (2016). We adjust the required sample size by a variance inflation factor (List et al. 2011: 451) of $1 + (t-1)\rho$, where $t = 16$ is the number of rounds. The factor $\rho = \frac{var\,(u_i)}{var\,(u_i)+var\,(e_{it})}$ denotes the extent of intra-subject correlation, where $var\,(u_i)$ captures between-subject variation and $var\,(e_{it})$ captures within-subject variation. We estimate $\rho = 0.12$, so the required sample should be corrected for the variance inflation factor of $1 + (16-1)0.12 = 2.8$. Dividing the number of observations required from the standard power test by the 16 observations and multiplying by this factor, we obtain a required sample size of $\frac{332}{16}2.8 = 58$ Auditors for each treatment. We therefore aimed at collecting 60 observations per role and treatment, as preregistered in AsPredicted #80975.

References

Moffatt, P. G. (2016) Experimetrics. Econometrics for experimental economics. London, New York: Macmillan Education Palgrave.

List, J. A., S: Sadoff, and M. Wagner (2011) So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics* 14 (4): 439–457