International Review of Law & Economics 78 (2024) 106189

Contents lists available at ScienceDirect

International Review of Law & Economics

journal homepage: www.elsevier.com/locate/irle

Is transparency a blessing or a curse? An experimental horse race between accountability and extortionary corruption

Christoph Engel [a,*], Eyal Zamir [b]

[a] Max Planck Institute for Research on Collective Goods, Bonn 53113, Germany
[b] Faculty of Law, Hebrew University, Jerusalem 9190501, Israel

ARTICLE INFO

JEL Classifications:
C91
D02
D63
D73
H11
H26
H83

Keywords:
transparency
anonymity
experiment
sequential game
tax evasion
extortionary corruption

ABSTRACT

If it is disclosed to a citizen which public official handles her case, this creates accountability. If the official abuses her authority, the citizen can report this misconduct to higher authority, which can intervene. But transparency also makes it possible for a citizen to pressure an official to decide in her favor. We model this interaction as a sequential game, and define which behavioral effects are required for either effect to dominate. We test the game experimentally. Within the parameters of our experiment, transparency clearly trumps anonymity. If the abuse of sovereign authority risks going unchecked, the occasional retaliation against dutiful officials is, on balance, the smaller social cost.

1. Introduction

As Justice Brandeis famously claimed, sunlight is the best disinfectant (Brandeis, 1914, 92). Transparency yields accountability (see e.g. Bovens, 2005, Hood, 2010). However, transparency is not a panacea. Anonymity, in some instances, may facilitate unpopular decisions, help attract higher campaign donations (Gardner, 2010), or enable referees to be critical (Campanario and Juan Miguel, 1998a, Campanario and Juan Miguel, 1998b). In this paper, we focus on what is arguably the most important effect: anonymity shields the decision-maker from pressure and temptation. This, for example, explains why the courts have the power to conceal the identity of jury members[1]; why the identity of the members of the Index Committee of Standard & Poors 500 is kept confidential[2]; and why the World Health Organization only disclosed the composition of the committee in charge of responses to the 2009 influenza outbreak after they had completed their work (Zamir and Engel, 2021, 1076 f.).[3]

Another illustration is an association having bylaws that require its members to pay membership fees. These fees are received by a particular official, who is also tasked with auditing payments and, if fees are withheld, taking action. If the identity of the official is known, he or she may be afraid that a member unwilling to pay will go after her, for instance by making a false accusation, if she takes action. Yet the official has first-hand information about payments as they arrive, and may exploit this knowledge to embezzle some of the dues. If he is accountable to members, individual members may spot signs of his misconduct, and report them to the association's board.

The analysis and findings of this paper do not only matter if the identity of a public official is kept confidential. They are also relevant to other means used to shield decision-makers from external pressure and

* Corresponding author.
E-mail address: engel@coll.mpg.de (C. Engel).

[1] Leading cases are *United States v. Barnes*, 604 F.2d 121, 140–41 (2d Cir. 1979); *United States v. Ross*, 33 F.3d 1507, 1519–22 (11th Cir. 1994).

[2] https://www.wsj.com/articles/david-m-blitzer-stock-pickerbehind-the-s-p-500–1495186202 [https://perma.cc/VAX8-LQN5].

[3] World Health Organization, The International Response to the Influenza Pandemic: WHO Responds to the Critics (2010), https://www.who.int/csr/disease/swineflu/notes/briefing_20100610/en/.

https://doi.org/10.1016/j.irle.2024.106189
Received 1 July 2023; Received in revised form 29 January 2024; Accepted 21 February 2024
Available online 29 February 2024
0144-8188/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

temptation. For example, in some instances the decision itself may be concealed. Thus, whenever we refer to "anonymity," this should be understood as including other instances in which a decision is not transparent.

We argue that, conceptually, transparency is a double-edged sword. It may deter agents—and public servants in particular—from abusing the powers they are entrusted with. But transparency also makes it possible for those who are subject to public intervention to exert pressure on the public servant, in an effort to sway her decision in their favor. We modeled this situation as a sequential game, and defined the parameters necessary for one of the two effects to dominate, with an eye to behavioral effects that are required for either effect. We then translated this game into the design of an experiment, which we ran on the platform *Prolific*, with a sizeable (N = 558) sample from the general public. To the best of our knowledge, this tradeoff has never before been examined empirically, let alone experimentally.

The participants in our experiment were randomly assigned the roles of citizens or public officials, and each citizen was paired with an official. The citizen decides whether to truthfully report her income and pay a tax—or to refrain from reporting her income, and evade the tax. If there is tax evasion, the public official may either take action (at a cost to herself) to have the citizen penalized—or not. If the official does take action, in the *Transparency* condition the citizen may retaliate (at some cost to herself)—thereby inflicting a loss on the official. If the taxes are paid, the official can embezzle the money. In the *Transparency* condition, the citizen can report the embezzlement to the authorities. Such reporting would be costly for the citizen—but much less so than the penalty meted out against the official. In the *Anonymity* condition, the citizen cannot report the errant official to the authorities.

Experiments are tools for causal inference. They are designed such that alternative explanations are ruled out. Participants are randomly assigned to conditions. If there is a significant difference in outcomes between conditions, it must result from the treatment manipulation. The inference is particularly credible if all choices are incentivized. In our experiment, we follow this tradition of experimental (law and) economics. One of us has shown in a related experiment that punishing a briber citizen less harshly than the corrupt official who accepts a bribe provides the citizen with an effective means to enforce the corrupt deal (Engel et al., 2016): if they accept the bribe, the officials are more likely to grant the favor if punishment is asymmetric. The present experiment exploits the same logic and investigates whether the prospect of retaliation also deters law enforcement.

Yet we readily acknowledge the limitations inherent in our choice of method. On the one hand, we abstract from many features of the real-life situations that we want to help understand. In this respect, higher internal validity comes at the usual price: one may discuss to which degree real-life analogues correspond to the stylized facts of the experiment. In our experiment, the citizen is the only person who can monitor the official; in real life, public officials are often also supervised from within government (but the citizen with whom they interact may have easier and better access to information). In the experiment, both the public official and the citizen incur a personal, monetary cost if they intervene; in real life, for the official the cost typically only comes in the form of hassle, and for the citizen it comes in the form of the risk that other officials (to whom the citizen reports) try to protect their colleague, and exert pressure on the citizen to withdraw the complaint.

On the other hand, we have disclosed to our participants that we are interested in a situation in which both tax evasion and corruption are technically possible. Hence our experiment has been framed. It has often been shown that frames may have a pronounced effect on the behavior of experimental participants (Tversky and Kahneman, 1981, Tversky and Kahneman, 1986, Tversky and Simonson, 1993, Engel, David, 2014, Engel and Reuben, 2015). We mainly have done so for an experimental reason. We model the social problem that we want to understand as a sequential game. Such games are not easy for participants to understand. By explaining to them which social conflict the game tree is meant to capture, we want to make sure that our data do not suffer from lack of understanding. This procedure is not uncommon in experimental economics. For instance, it is standard in market experiments to inform participants that firms are competing over profit (and not only instructing them about a naked incentive structure) (for detail see the meta study by Engel, 2015).

A further limitation is inherent in the experimental method. In a model, one can abstractly define the range for which a theoretical claim holds. But experimental participants need a fully defined situation. Hence experiments are parameterized. Now behaviorally, parameters may well matter. An effect predicted by theory may empirically only obtain if the difference between outcomes is sufficiently pronounced. Ultimately, if one is concerned about this possibility, one must repeat the (otherwise identical) experiment with different parameters. Doing this is beyond the scope of the present project. In our experiment, both the citizen and the official may be tempted by a relatively large, potential illicit gain (from tax evasion; from embezzlement) and both incur a considerably smaller cost of intervention (against tax evasion; against embezzlement). The absolute and relative sizes of these gains and costs vary across real-life situations, and may be more or less pronounced in some contexts.

Ultimately, lab experiments on legal issues always require a leap of faith. What one studies is only analogous to what one wants to understand. But the reader may be sure which cause has had which effect. In the concluding session, we will delve deeper into potential differences between the design of the experiment and real-life applications. Ultimately, all we can contribute to the policy debate is isolating one causal pathway. Yet we believe that understanding this causal chain is an important input into the policy choice between revealing and concealing the identity of the public official who handles a case (or otherwise making decisions more or less transparent): does the risk of retaliation against the legitimate exercise of sovereign powers against a law-breaking citizen loom larger than the risk that a public official abuses her powers if not afraid of citizen scrutiny, or the other way round?

We found a straightforward transparency effect. When shielded from retaliation, public officials are also shielded from citizen scrutiny. 66% of public officials in the experiment exploited the resulting absence of deterrence and embezzled the (experimental) tax proceeds. By contrast, when the interaction was transparent—such that the citizen who had paid the taxes could report the embezzlement, which the official was then penalized for—the incidence of embezzlement dropped to 33%. On the other hand, only 17% of all citizens in the experiment opted to retaliate against an official who had penalized them for tax evasion. Hence in the experiment, the social cost of transparency (the risk of retaliation) was much smaller than the social cost of anonymity (the risk that an unsupervised public official abuses her position to enrich herself).

In the following section, we relate our paper to the empirical literature on tax evasion and corruption. Section 3 introduces the experiment's design and predicts treatment effects. Section 4 reports results. Section 5 concludes with a discussion.

2. Literature review

The behavioral literature on transparency as an element of institutional design is rather small (we review this literature in Zamir and Engel, 2021). But the two building blocks of the setting into which we introduce transparency have attracted a lot of scientific interest: tax evasion and corruption. Both are easy to implement in the lab, and have been used *pars pro toto* for the broader question of regulatory enforcement (for experimental studies on this broader topic see Cason et al., 2006, Sundström, 2012, Telle, 2013, Blundell, 2020).

Corruption has not only attracted the interest of (economic) theorists (a classic is Allingham and Sandmo, 1972), but has also been widely studied empirically (for summary accounts see Andreoni et al. 1998, Blackwell, 2007, Kirchler, 2007, Torgler, 2007, Alm, 2012, Pickhardt

and Prinz, 2014). For decades it has been investigated how likely experimental participants are to evade taxes (Baldry, 1986, Swenson, 1988, Alm, 1991, Alm et al., 1992b), and under which conditions (Anderhub et al. 2001, Bayer and Sutter, 2009, Balafoutas et al. 2015). Multiple field experiments have attempted to verify the lab results under real world conditions (Blumenthal et al. 2001, Torgler, 2004, Coleman, 2007, Wan, 2010, Kleven et al., 2011, Sælen and Kallbekken, 2011, Ariel and Barak, 2012, Battiston and Gamba, 2013, McKee and Vossler, 2013, Gangl, Torgler et al. 2014, Castro and Scartascini, 2015, Dwenger and Kleven et al. 2016, Filer and Hanousek et al. 2016).

Much of the academic attention has been on tax morale (Alm et al. 1992a, Alm et al. 1995, Cummings et al. 2009, Luttmer and Singhal, 2014, Casal et al. 2022), explaining it with intrinsic morality (Hanno and Violette, 1996, Bosco and Mittone, 1997, Bobek and Hatfield, 2003), psychological cost (Coricelli et al. 2010, Coricelli et al. 2014, Dulleck et al. 2016, Enachescu et al. 2019), rule following (Engel et al. 2020), obedience (Cadsby et al. 2006), civic virtue (McGraw et al., 1991), respect for the law (Blaufus et al. 2016, Engel, Mittone et al. 2024), social norms (Hanno and Violette, 1996, Bobek et al., 2013, Abraham et al. 2017, Doerrenberg and Peichl, 2017), education (Rodriguez-Justicia and Theilen, 2018) and shaming (Casagrande et al., 2015, Perez-Truglia and Troiano, 2015).

Another series of experiments has investigated the social dimension of tax compliance (Christian and Alm, 2014), highlighting the importance of social preferences (Ackert et al., 2007), perceived fairness (Bordignon, 1993, Jimenez and Iyer, 2016), reciprocity (Bazart and Bonein, 2014, Doerrenberg and Peichl, 2017), rivalistic preferences like a competitive spirit (Casagrande et al., 2015) or spite (Cullis, Jones et al. 2012), conditional cooperation (Traxler, 2010) and sensitivity towards procedural fairness (Niesiobędzka and Kołodziej, 2019). Other experiments have focused on risk preferences (Magessi et al., 2015), and on cognitive effects (McCaffery, 2000), like framing (Fochmann and Wolf, 2019) or salience (Chetty et al. 2009).

A further strand of the literature investigates the relationship between perceived legitimacy of the tax code and compliance (Engel et al. 2024), in terms of both the purpose for which the public budget shall be used (Falkinger, 1995, Doerrenberg, 2015), including redistribution (Doerrenberg and Duncan, 2014), and participation in the generation of rules (Güth and Sausgruber, 2008, Wahl et al., 2010).

The empirical literature has been sensitive to the possibility of cultural differences and has compared results from multiple countries (Batrancea et al. 2019), comparing the United Kingdom and Italy (Zhang et al. 2016), Australia, Singapore and the United States (Bobek et al., 2007), Belgium, France and the Netherlands (Lefebvre et al. 2015), and studying samples in China (Kao, 2016), the Ukraine (Bilotkach, 2005), and Costa Rica (Torgler, 2003).

Most relevant for our paper are experiments that have tested institutional interventions. Some studies have introduced nudges, like prefilling the tax return (Kotakorpi and Laamanen, 2016), increasing the salience of audit (Bott et al., 2020), or public scrutiny (Garcia et al. 2020). One study has introduced rewards (Fochmann and Eike, 2016). But the majority of these studies have focused on deterrence, increasing the probability of audit (Spicer et al., 1982, Slemrod et al. 2001), strategically making audit less foreseeable (Tan and Yim, 2014), or comparing the effect of detection probability vs. expected severity of punishment (Jackson and Jones, 1985).

Our paper contributes to deterrence research and adds a hitherto neglected dimension. In previous studies, deterrence has been automatic, and perfectly predictable. By contrast, in our experiment the probability of enforcement is itself a function of choices made by human participants, for whom enforcement comes at a (relatively small) cost. We believe this manipulation not only casts light on the black box of decision making within the tax authorities. More importantly even, from a policy perspective, we see how decisions to be made by tax officials are anticipated, and potentially influenced, by tax payers.

2.1. Corruption

Our experiment essentially compares the risk of tax evasion with the risk of corruption. Corruption has also been the topic of a lively experimental literature (for summary accounts, see Abbink and Serra, 2012, Lambsdorff, 2012, Bobkova and Egbert, 2013, Incerti and Trevor, 2020), also in this journal (Lambsdorff and Frank, 2011), and including an earlier contribution by one of us (Engel et al., 2016).

For instance, lab experiments have focused on reciprocity between bribers and public officials (Abbink et al. 2002), other-regarding preferences (García-Gallego et al. 2020), social norms (Banerjee, 2016), the effects of framing (Barr and Serra, 2009), whether the same individuals interact repeatedly (Balafoutas, 2011), whether monitors are appointed or elected (Barr et al., 2009), and whether public officials compete with each other (Ryvkin and Serra, 2020).

As experimenters have worried about external validity (Armantier et al. 2012), this literature has been complemented by a fair amount of field experiments in Burkina Faso (Armantier and Boly, 2011), Canada (Armantier and Boly, 2013), Costa Rica (Corbacho et al., 2016), Indonesia (Olken, 2007), Liberia (Beekman et al. 2014) and Afghanistan (Callen and Long, 2015). There is pronounced cross-cultural variation (Cameron et al. 2005, Alatas et al. 2009b, Alatas et al. 2009a, Barr and Serra, 2010, Zhang, 2015, Salmon and Serra, 2017, Zhang, 2018).

Other papers have manipulated institutional interventions: is accountability organized bottom up or top down (Serra and Danila, 2012)? Are bribers or bribees rewarded for self-reporting (Abbink and Wu, 2017)? For our purposes, the most relevant papers are the ones that have tested the effect of increased transparency. García-Gallego et al. (2020) show that, indeed, the presence of independent observers reduces the incidence of corruption, as does the presence of passive participants whose payoff is reduced by corruption. But combining both interventions does not further decrease corruption. In Khadjavi et al. (2017), mere transparency was not instrumental. It must be combined with sufficient punishment power to prevent an official from engaging in embezzlement.

Peisakhin and Pinto (2010) show that, in India, freedom of information legislation has empowered even the poorest of the poor, by enabling them to make corruption publicly known. In a meta-study, Incerti and Trevor, 2020 shows that making corruption transparent has a large effect when surveying participants, but almost no effect in field experiments. Mansour et al. (2021) compare two technologies for citizen scrutiny of a potentially corrupt official. In one condition, experimental citizens can only voice dissatisfaction. In the other condition, they can limit the official's term. Both interventions turn out to have a dampening effect. But voice becomes counterproductive if removing the official from the office is possible. Apparently, in this condition, officials interpret the expression of dissatisfaction as irrelevant cheap talk.

All of these studies show an effect of citizen empowerment, but they are not allowing for a potential downside, as in our experiment. A notable exception is Chong et al. (2015), where learning about the degree of corruption (in local elections in Mexico) did decrease voter turnout. But that downside is very different from the one we investigate.

3. Design and predictions

We were interested in the effect of concealing the identity of a public official from those affected by her decisions (or of making her decisions otherwise unobservable). We considered the possibility that both the affected individual and the public official were selfish, and prepared to violate a rule that constrained them. If the public official observed a violation, she had the power to report it and trigger a sanction; similarly, the citizen could report violations by the public official—but only if the latter's behavior was *transparent*. Finally, we made it possible for the citizen to retaliate against a public official who had penalized her. *Anonymity* simultaneously closes two possible channels: the citizen can no

longer report the public official for breaching of the public's trust by embezzling the tax proceeds, and the public official is no longer subject to the threat of retaliation. Thus, anonymity potentially has both a normatively undesirable effect (the public official avoids scrutiny) and a normatively desirable one (the public official is safe from retaliation when she takes action against tax evasion). The purpose of the experiment was to examine which effect was stronger.

If all participants are exclusively interested in material payoff, expect that the random person with whom they interact is also exclusively motivated by material payoff, if they expect the first person to be thus motivated (and hence higher order beliefs are in line with first order beliefs), and possess the necessary cognitive abilities— i.e. if common knowledge of rationality can be assumed—the interaction constitutes a sequential game. This game can be solved by backward induction. In the appendix, we formally derive the solution, and hence the equilibrium of the game. We do so for generic parameters. We thus define the relationship between parameters that are required for the respective predictions to hold. In a second step, we relax the rationality assumption, and formally define deviations from common knowledge of rationality that would be necessary to change the predictions. However, in the experiment we cannot have participants decide for a rich matrix of parameters. Rather, we zero in on a set of parameters that make sense, given the policy problem that motivates the experiment. Fig. 1 graphically defines the design of the experiment in the form of a game tree, and with the parameters that we have implemented in the lab. The participants have initially seen this exact figure (which in subsequent steps of the instructions, we have broken down into easily understandable subfigures).

As shown in Fig. 1, pertaining to the *Transparency* condition, the first decision is made by the citizen, who either declares her income (in which case she pays a tax of 5 out of her initial endowment of 15), or refrains from doing so (thereby keeping the entire endowment of 15 for herself). If the citizen declares her income and pays the tax of 5, the official can either embezzle the proceeds (thereby increasing her initial endowment of 10–15), or refrain from doing so and retain her initial endowment of 10. If the citizen evades the tax, the official can either penalize her (at a cost of 2 to herself—leaving her with 8 of her initial endowment of 10), or refrain from doing so. If the official takes such action, the citizen is fined 10, and is left with 5. If the official does not take such action, the official is left with her initial endowment of 10, and the citizen with her initial endowment of 15.

In the *Anonymity* condition (represented in Fig. 2), these possibilities exhaust all possible courses of action by the citizen and the official. Conversely, in the *Transparency* condition, the citizen can react to the official's actions. Thus, if the citizen declares her income (consequently reducing her endowment from 15 to 10), and the official does not embezzle the proceeds, both the citizen and the official end up with 10, and the citizen has no recourse of action. However, if the citizen declares her income and the official embezzles the tax proceeds, the citizen must decide whether to report the official's transgression, at a cost of 2 to herself—in which case she ends up with 8, and the official is fined 10, and ends up with 5. If the citizen does not declare her income and the official takes no action, no reaction by the citizen is possible. In that case, the citizen retains her initial endowment of 15, and the official keeps her initial endowment of 10.

Finally, if the citizen does not report her income and the official penalizes her for it, in the *Transparency* condition the citizen faces a choice whether to retaliate (at a cost of 2 to herself), or not. If she retaliates, both parties end up with 3: the citizen, who initially had 15, has been fined 10 for evading the tax and bears the additional retaliation cost of 2; the official, who initially had 10, bears the cost of her action (2), then loses an additional 5 due to the citizen's retaliation. If the citizen chooses not to retaliate, she ends up with 5 (as a result of being fined for not reporting her income), and the official ends up with 8 (her initial endowment of 10, minus the costs of acting in response to the tax evasion).

With common knowledge of standard preferences, both the citizen and the public official are expected to maximize their respective profit. Under these assumptions, anonymity is immaterial: reporting and threatening the public official would both be costly and risky. Therefore, an individual exclusively interested in personal profit is not expected to resort to either option. Hence with common knowledge of standard preferences, the citizen is not deterred from evading taxes, and the official does not expect to have the opportunity to embezzle the tax proceeds. This gives us our null hypothesis:

H_0 (Common Knowledge of Standard Preferences):

a) The citizen evades taxes
b) The official is indifferent as to the conditional opportunity to embezzle the tax proceeds, or not
c) The official does not penalize the citizen for tax evasion
d) The citizen does not penalize the official for embezzlement
e) The citizen does not conditionally decide to retaliate, if penalized by the official
f) There are no treatment effects.

However, reporting and retaliation may be interpreted as punishment. Specifically, revenge can be interpreted as second party punishment (Fehr and Gächter, 2000, Fehr and Gächter, 2002): the citizen resents being punished, and reacts by counter-punishment (Nikiforakis, 2008). This type of punishment has, in fact, been shown experimentally to stabilize corrupt deals that are legally unenforceable, and incentive-incompatible (Engel et al., 2016). Arguably, the official may also engage in second-party punishment: if the citizen evades taxes, the official is denied the opportunity to embezzle. If the official had intended to embezzle the tax proceeds, threatening the citizen with punishment may be interpreted as an act of extortionary corruption. Assuming that the citizen anticipates that, otherwise, she will be fined, the citizen feels forced to declare her income, and to provide the public official with the opportunity to embezzle the tax she has paid. Actually, if this composition of behavioral effects is critical, anticipating that public officials will want to graft some of the money would make the threat with prosecution credible, and hence would deter tax evasion.[4] If there is only second-party punishment, and this is anticipated, treatment matters. In the *Anonymity* condition, only the official can punish, while in the *Transparency* condition, both players can. This gives us our first alternative hypothesis:

H_1 (Second-Party Punishment):

a) In the *Anonymity* condition—

i. The citizen pays taxes
ii. The official embezzles the tax proceeds
iii. The official penalizes the citizen for tax evasion

b) In the Transparency condition—
i. The citizen does not pay taxes
ii. The official is indifferent to the conditional opportunity to embezzle the tax proceeds, or not
iii. The official does not penalize the citizen for tax evasion
iv. The citizen does not penalize the official for embezzlement
v. The citizen retaliates if penalized by the official

In the experimental literature, third-party punishment has also been documented. When person C observes person A mistreating person B, C punishes A for the mistreatment (Fehr and Fischbacher, 2004, Rockenbach and Milinski, 2006, Almenberg et al. 2011, Leibbrandt and López-Pérez, 2012, Balafoutas et al. 2014, Lergetporer et al. 2014, Nikiforakis and Mitchell, 2014). This also

[4] We are grateful to an anonymous referee for pointing this implication out to us.

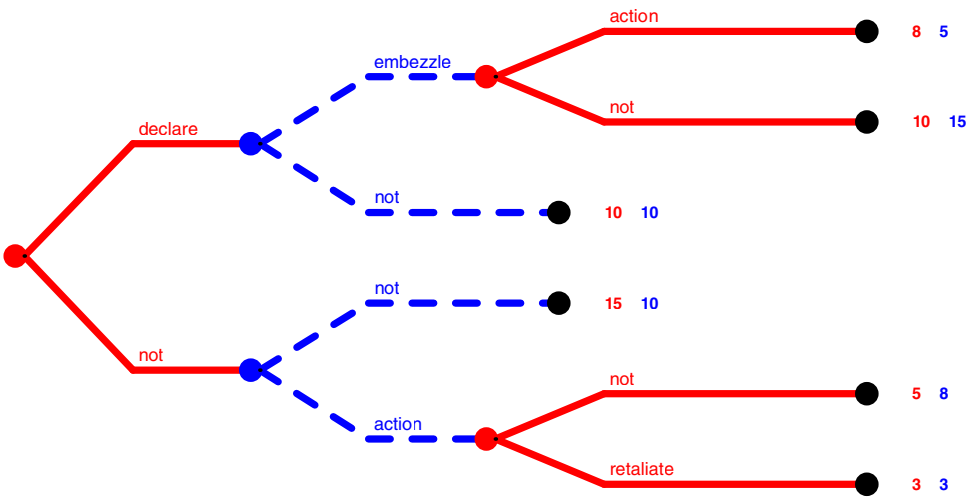

Fig. 1. Game Tree Assuming Common Knowledge of Standard Preferences: Transparency Condition, citizen's moves and final payoffs in red / solid, official's moves and final payoffs in blue / dashed.

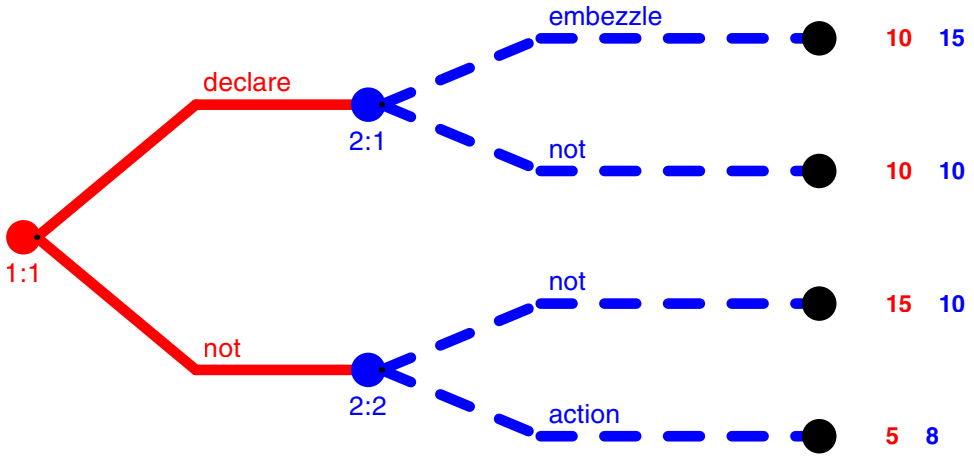

Fig. 2. Game Tree Assuming Common Knowledge of Standard Preferences: Anonymity Condition, citizen's moves and final payoffs in red, official's moves and final payoffs in blue.

holds true if one participant is assigned the role of an authority with the power to discipline a group: in experiments, a large majority of people in authority do use their power—effectively stabilizing cooperation in a dilemma game, even if this involves a significant cost for themselves (Engel and Zhurakhovska, 2017).

In the present experiment, third-party punishment was required for the citizen to punish an official who had embezzled tax proceeds.[5] Third-party punishment was also required if the official punished the citizen for tax evasion while intending to embezzle the tax proceeds. If there is only third-party punishment and no second-party punishment, we have hypothesis H_2:

H_2 (Third-party Punishment):

a) In the *Anonymity* condition—

i. The citizen pays taxes
ii. The official embezzles the tax proceeds
iii. The official penalizes the citizen for tax evasion

b) In the Transparency condition—

i. The citizen does pay taxes
ii. The official does not embezzle the tax proceeds
iii. The official penalizes the citizen for tax evasion
iv. The citizen penalizes the official for embezzlement
v. The citizen does not retaliate if penalized by the official

Our experiment was prompted by the expectation that there would be both second-party punishment *and* third-party punishment—and, correspondingly, that at least some citizens would be tempted to evade taxes, and at least some public officials would be tempted to embezzle the tax proceeds. We therefore expected conflicting behavioral effects. In Appendix A, we define which parameters are required for *Anonymity* to outperform *Transparency*. In a nutshell, *Anonymity* is more effective if (a) many officials enforce taxation and (b) few officials are dishonest (intuitively, under these conditions there is less need for transparency to facilitate the external monitoring of officials' behavior). Conversely, *Transparency* is more effective if (a) many officials enforce taxation, or many citizens are honest (so that the stipulated taxes are collected), (b) many officials are dishonest (thus making external monitoring more essential), (c) many citizens intervene against embezzlement (otherwise transparency does not improve the situation) and (d) few citizens retaliate when penalized (thus diminishing the concern about the adverse effect of transparency as facilitating retaliation against officials

[5] In our experiment, the citizen is not a completely disinterested third party (since the official embezzles the tax proceeds that the citizen has paid)—but when reacting to an embezzlement, she is doing so in response to an action that harms the state (the experimenter), not herself.

who dutifully enforce tax law). This provides our two main hypotheses:[6]

H_3 (Anonymity Outperforms Transparency):

a) In the *Anonymity* condition—
i. Most citizens pay taxes
ii. Most officials do not embezzle the taxes paid by the citizen
iii. Most officials penalize tax evasion

b) in the Transparency condition—
i. Most citizens do not pay taxes
ii. Few officials penalize tax evasion
iii. Many citizens retaliate if penalized by the official

Thus, if *Anonymity* outperforms *Transparency*, tax evasion is the first-order problem. If, by contrast, embezzlement is the first-order problem, we expect *Transparency* to outperform *Anonymity*, and predict:

H_4: (Transparency Outperforms Anonymity)

a) In the *Anonymity* condition—
most officials embezzle the tax proceeds
b) In the Transparency condition—
i. Many citizens pay taxes
ii. Most officials do not embezzle the tax proceeds
iii. Many officials penalize tax evasion
iv. Many citizens penalize embezzlement
v. Few citizens retaliate if penalized by the official

It is, of course, possible that the effect of *Transparency* vs. *Anonymity* on officials and citizens varies from one person to another. In that case, the governance problem that prompted this inquiry is more acute, as the policymaker does not know which individuals will later come under the purview of any given rule. Accordingly, if a given rule's effect varies with circumstances, the policymaker should get the overall effect right. The rule cannot always be ideal, but it should at least be appropriate most of the time. Therefore, while policymakers are not constrained by the possibility of behavioral heterogeneity per se, they do need a reliable estimate of the prevalence of competing behavioral effects. The present study sought to produce just such an estimate.

Behavioral heterogeneity exacerbates the problem for individual citizens and public officials. It leads to uncertainty about the actions of one's experimental counterpart. This uncertainty is substantial, because participants lack not only information about what their individual counterpart is going to do, but also reliable information as to the probabilities that the official might penalize tax evasion, the citizen might retaliate in response, the citizen will report the official for embezzlement, and the citizen will take revenge if sanctioned. Participants therefore face ambiguity (Heath and Tversky, 1991, Weber and Camerer, 1992). They must work with subjective estimates of these probabilities (Savage, 1954), and evaluate these uncertain effects with their own degree of aversion to ambiguity (Machina and Siniscalchi, 2014). In Appendix A, we explain the ways in which this uncertainty moderates the effects.

We reflected this uncertainty in the experimental design by eliciting and incentivizing beliefs about all five decision options by the respective counterpart.[7] In order not to alert participants to behavioral risks on the part of their random interaction partners, we elicited participants' beliefs only after they had made all of their choices. Specifically, we learned from citizens what proportion of the officials in their experiment session they believed would embezzle the tax proceeds (if the tax was paid), and how many of the officials would penalize citizens for evading taxes. In the *Anonymity* condition, we learned from officials how many citizens in their session they believed would evade taxes. In the *Transparency* condition, we further learned how many citizens they believed would report them if they embezzled tax proceeds, and how many citizens they believed would retaliate if they (the officials) penalized them for tax evasion.

The experiment allows for sanctions. We can therefore study to which degree the expectation that the experimental counterpart engages in second- or third-party punishment deters tax evasion, or the embezzlement of taxes. The deterrent effect, however, presupposes that (a) citizens are willing to incur the cost of punishment although this has no material benefit for them and (b) officials anticipate this reaction. Moreover, a citizen may be expected to pay taxes irrespective of the threat with enforcement. In the interest of controlling for such moral compunctions, in a post-experimental survey we asked how blameworthy they considered a citizen who evaded taxes, or an official who embezzled tax proceeds. Finally, we collected demographic information.

To generate full data, we used the *strategy method* (Selten, 1967)—namely, participants made conditional choices based on the corresponding actions by the other member of their random group of two (the citizen evading taxes; the official embezzling the tax proceeds; the official imposing a penalty for tax evasion).[8]

We preregistered the experiment on the Open Science Framework,[9] and programmed it with the oTree software application (Chen et al. 2016). Since sequential games may be difficult for the uninitiated, we unfolded the game gradually. In a series of slides, we explained each final node of the game tree, and the individual path leading to it (see Appendix C). In these explanations, we always coupled the picture of the game tree (with the final node in question highlighted) with a verbal explanation. Participants were then asked a series of comprehension questions. As we were interested in choices that depend on the assigned role, we only had one independent observation per group of two. We had preregistered that we need 100 observations (groups) per condition, to be able to establish an effect of size Cohen's d =.4, assuming the conventional values of α =.05 and β =.2. To have even more statistical power, we actually collected 142 independent observations in the *Anonymity* condition, and 137 in the *Transparency* condition. This enables us to establish an effect of Cohen's d =.345. We ran the experiment on the Prolific platform.[10] Table 1 shows the composition of the sample and earnings.

4. Results

The main results are presented in Fig. 3, revealing a straightforward transparency effect. In the *Anonymity* condition, 66% of all officials decided to embezzle the tax proceeds if the citizen in their group declared her income. In the *Transparency* condition, this proportion dropped to 33%. Correspondingly, in the *Transparency* condition, 40% of all citizens decided to report an official who embezzled the tax proceeds. Conclusion: In the experiment, transparency dramatically reduces corruption.

Moreover, only 17% of citizens decided to retaliate against an official who penalized them for tax evasion. The main concern that motivated our inquiry—namely, the expectation that anonymity might be instrumental in shielding the official from pressure and temptation—was not supported by the experimental data. Similarly, the decision of officials to intervene against tax evasion was only marginally reduced in the *Transparency* condition—from 36% to 34%. The incidence of tax evasion rose only marginally in the *Transparency* condition—from 23% to 26%.

[6] We preregistered only these main hypotheses.
[7] For details of the incentive scheme, see instructions in the Appendix.
[8] The one exception is the decision by the citizen to intervene against embezzlement of the tax proceeds by the official, when the citizen herself has not reported her income. We felt it implausible and artificial to ask a citizen who has decided to evade taxes (so the official has nothing to embezzle) whether she would penalize the official in the counterfactual situation that she actually had paid taxes and the official had embezzled them.
[9] https://osf.io/zu7pv/?view_only=3356f3da680446dca8873aa410c5a6f1.
[10] https://www.prolific.co.

Table 1
Sample.

Treatment	Role	N	Payoff	Female	age	UK	High school	Under grad	Grad	Full- time job	Native English
anonymity	citizen	142	£ 7.92	40.1%	37.0	57.7%	24.6%	42.3%	19.7%	59.9%	80.3%
	official	142	£ 8.72	39.4%	38.1	63.4%	22.5%	44.4%	19.0%	47.9%	79.6%
transparency	citizen	137	£ 7.71	48.2%	36.4	65.7%	34.3%	38.0%	19.7%	64.2%	84.7%
	official	137	£ 8.15	50.4%	38.6	67.9%	31.4%	43.1%	20.4%	46.7%	89.1%

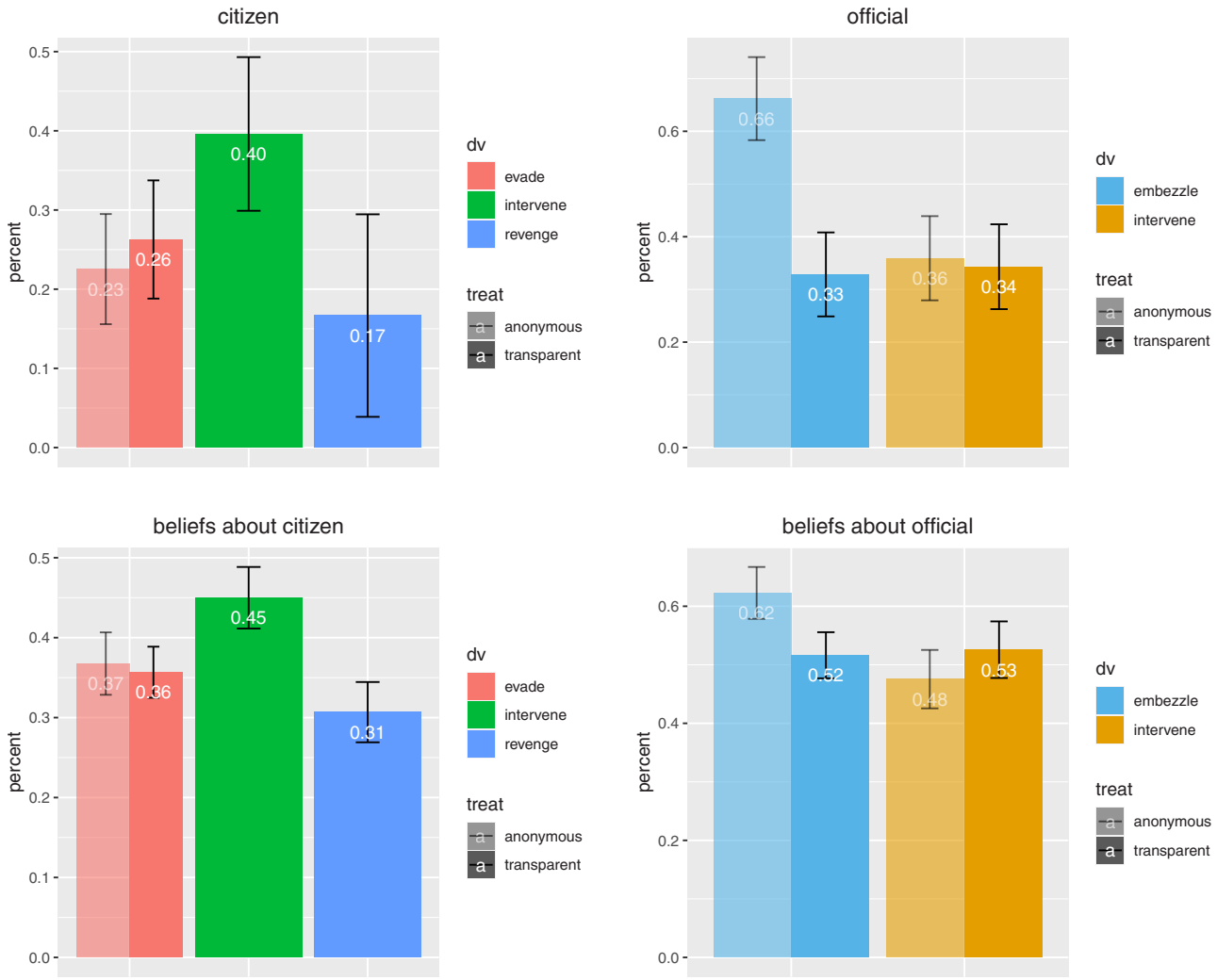

Fig. 3. Choices and Beliefs. upper left panel: choices of participants in citizen role; evade: binary decision not to report income; intervene: binary decision to intervene, should official embezzle paid tax; revenge: binary decision to inflict harm on official if she has reported tax evasion; upper right panel: choices of participant in role of official; embezzle: binary decision to keep paid tax; intervene: binary decision to intervene, should citizen withhold taxes; lower left panel: beliefs of official about citizen choices; lower right panel: beliefs of citizen about official's choices; lighter color: anonymous condition; error bars: 95% confidence interval.

Additionally, the two lower panels of Fig. 3 reveal a pronounced mismatch between choices and beliefs. The difference is most striking in the case of retaliation: officials believed that 31% of all citizens would take revenge, while in reality only 17% did. There was also a pronounced difference between actual embezzlement decisions and beliefs. Citizens had a reasonably good sense of when embezzlement occurred in the absence of transparency (their belief was 62%, actual incidence is 66%). But they quite significantly underestimated how effective was the threat of punishment: while they thought that 52% of officials would still embezzle in the *Transparency* condition, in reality only 33% did. Conversely, citizens overestimated the likelihood that officials would penalize them if they (the citizens) evaded taxes (48% vs. 36% in the *Anonymity* condition; 53% vs. 34% in the *Transparency* condition).

We begin with statistical tests for treatment effects, as shown in Table 2.[11] We find a strong and highly significant effect on the decision of public officials to embezzle the tax proceeds (if there were any). In contrast, no significant treatment effect is found with regard to the decision of citizens to evade taxes, or to the decision of officials to penalize tax evasion. The same picture arises with regard to beliefs. But despite also being highly significant, the effect of *Transparency* on the belief that

[11] The regression equation is simply. $dv = \beta_0 + \beta_1\ treat + e$ We estimate linear effects as then coefficients can directly be interpreted as marginal effects.

Table 2
Treatment Effects on Choices and Beliefs.

	Choices			Beliefs		
	evade	embezzle	intervene against evasion	evade	embezzle	intervene against evasion
transparency	0.037	-0.334***	-0.016	-0.011	-0.107***	0.050
	(0.052)	(0.057)	(0.057)	(0.026)	(0.030)	(0.035)
cons	0.225***	0.662***	0.359***	0.368***	0.623***	0.475***
	(0.036)	(0.040)	(0.040)	(0.018)	(0.021)	(0.025)
N	279	279	279	273	272	272

OLS; evade, belief about embezzle and intervene against evasion: data from citizens; embezzle, intervene against evasion, belief about evade: data from officials; N on belief data differs from N on choice data due to a small number of online participants who left the experiment prematurely; standard errors in parenthesis *** p <.001, ** p <.01, * p <.05

the official would embezzle the tax proceeds is found to be much smaller (.107 versus .334) than the treatment effect on the officials' actual choices.

By design, two of the citizen's choices, and the official's beliefs about those choices, were only available in the *Transparency* condition. We cannot, therefore, test for treatment effects, but we can report results from one-sample t-tests.[12] With the help of these tests, we can define an effect of which size we can exclude, assuming the conventional α-level of.05. Put differently, we define a range within which the population effect may lie, given our findings. With regard to reports by citizens about embezzlement, we can exclude the possibility that the proportion is as low as 29% (p =.035), and as high as 50% (p =.036). For citizens retaliating for being penalized for tax evasion, we can exclude the possibility that the fraction is as low as 3% (p =.037) and as high as 30% (p =.041). With regard to the beliefs of officials about citizen intervention, we can exclude the possibility that the expected fraction is as low as 41% (p =.042) and as high as 49% (p =.042). With regard to the beliefs of officials about citizens retaliating for being penalized for tax evasion, we can exclude the possibility that the expected percentage is as low as 26% (p =.015), and as high as 35% (p =.025).

We also use t-tests to check whether there was a mismatch between choices and beliefs.[13] We find that this is always the case—with the exception of the belief about the incidence of embezzlement in the *Anonymity* condition (where we only find a marginally significant effect—p =.099). All other beliefs are significantly off (*evasion anonymity*: p <.001; *evasion transparency*: p <.001; *embezzlement transparency*: p <.001; *intervention against tax evasion anonymity*: p <.001; *intervention against tax evasion transparency*: p <.001; *intervention against embezzlement transparency*: p =.012; *revenge*: p <.001).

For the normative evaluation of our findings, it is important to know who stands to benefit in each instance. This assessment is easiest in the case of tax payment: while in the *Anonymity* condition, only 26.1% of the tax due is actually collected, in the *Transparency* condition this proportion jumps up to 48.9%. The public budget is the big loser. We can also calculate *efficiency*—i.e., the joint payouts to both players and the public. This turns out to be indistinguishable (25.6 ECU in the *Anonymity* condition vs. 25.3 ECU in the *Transparency* condition). But the distributional effect is clear: with *Anonymity*, the public loses, in favor of greedy citizens and public officials.

When they have the opportunity (i.e., in the *Transparency* condition), 40% of all citizens decide to call out embezzlement. This is unequivocal evidence of third-party punishment, and speaks against the null hypothesis $\mathbf{H_0}$, that assumes common knowledge of standard preferences. Descriptively, it is only 2 pp. less likely in the *Transparency* condition that the official intervenes against tax evasion. This suggests that this penalty is largely motivated by the intention to enforce the norm, rather than to extort money to be embezzled. With 17%, second-party punishment by (retaliating) citizens is rather infrequent. Both speaks against $\mathbf{H_1}$. Second-party punishment is, at least, not the primary effect. There is a substantial incidence of decisions in line with third-party punishment (40% reporting embezzlement; 36%/34% penalizing tax evasion). However the majority refrains from third-party punishment. A substantial proportion of citizens (23%/26%) evades taxes, and a substantial proportion of officials embezzles tax proceeds (66%/33%). We therefore also have no clear evidence in favor of $\mathbf{H_2}$—which was based on third-party punishment as the primary behavioral effect. The data suggests that we observe a heterogeneous population, a substantial percentage of whom are selfish (given the opportunity), and that the risk of a negative, retaliatory reaction appears to be sufficiently contained. Behavioral heterogeneity is the conceptual basis for competing hypotheses $\mathbf{H_3}$ and $\mathbf{H_4}$. Our evidence clearly supports the latter—namely, *Transparency* outperforms *Anonymity*.

In summary:

Result: The risk that a public official abuses her powers for personal gain looms larger than the risk that a citizen would force the official, by threatening her, to let her evade taxes.

For exploratory purposes, we also report results from our supplementary measures. In the concluding survey, we asked to what extent participants think it blameworthy if a citizen evades taxes, or an official embezzles tax proceeds. As Fig. 4 shows, on average, participants thought both actions to be blameworthy, but embezzlement more so than tax evasion. Differences between treatments and roles were negligible.

In Table 5 in Appendix A, we report regressions that check for correlations between choices and all other variables that we collected. Interestingly, beliefs are found to make little difference. The only effect that is significant at conventional levels is the belief of citizens about the likelihood of embezzlement by tax officials: the more they expect this to be the case, the less (not the more, as theory would suggest) likely they were to take action (in the *Transparency* condition). In contrast, in the case of the decision to evade taxes, we find the expected effect: the more citizens deem it likely that the official would penalize them for tax evasion, the less likely they were to do so. But the effect is only marginally significant (p =.095). The more officials found embezzlement to be blameworthy, the less they were likely to engage in it. The more they found tax evasion to be blameworthy, the more they were likely to penalize it (p =.084). Women were more likely to call out embezzlement—as were older participants. Those in full-time employment were more likely (as citizens) to evade taxes, and to embezzle tax proceeds (as tax officials).

5. Discussion

On conceptual grounds, many regulatory choices are indeterminate:

[12] Technically, we use a function over a plausible range of percentages, in increments of 1%, and report the lowest and the highest percentage that we can exclude, at the 5% significance level.

[13] Technically, we test beliefs against the observed mean of choices as a point prediction.

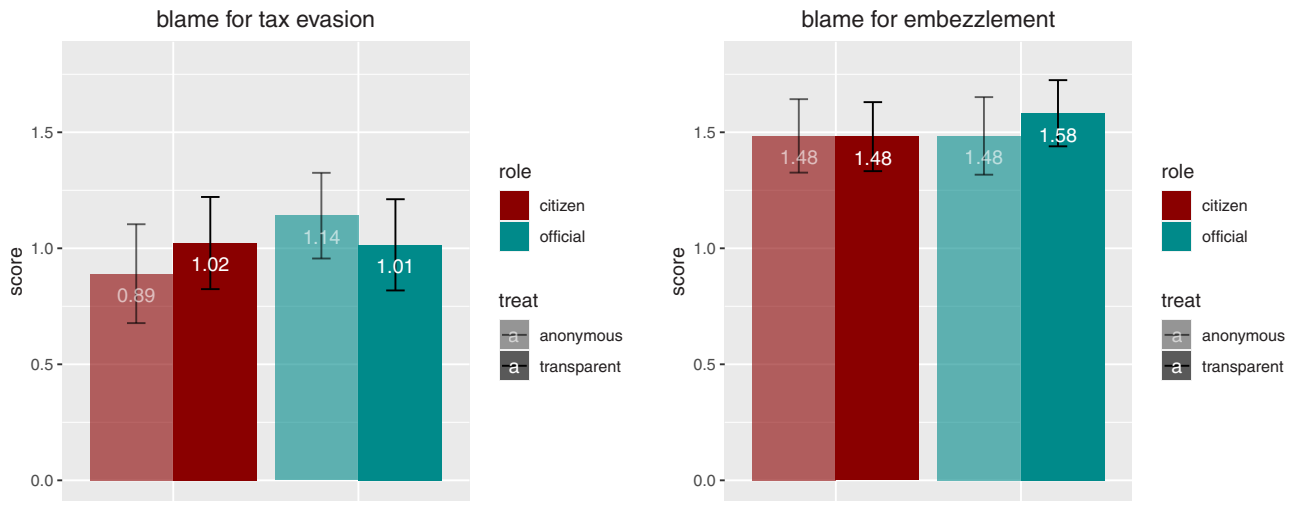

Fig. 4. Blameworthiness. left panel: stated blameworthiness of evading taxes, on a score from −2 to 2, role: citizen or official, treat: anonymous (lighter color) or transparent (darker color), error bars for 95% confidence interval.

consistent and plausible concerns may be advanced either way. The standard response is politics. The determining factors are the preferences of the majority of the population, the policies of the ruling political party, or some form of strategic interaction in Parliament or government. But for some political choices, a preference-based approach may not be suitable. This paper discusses one such choice: Is tax evasion worse than corruption? Our survey data on blameworthiness suggests otherwise. Yet the difference is not overwhelming: most participants find neither of them to be acceptable. That shifts the focus from normative judgments to effectiveness: if a technology is not available that reliably fixes both, policymakers may want to choose the intervention that is more effective.

If the identity of the public official who handles the case is disclosed and her decision is transparent, the relevant citizen may spot irregularities and trigger an investigation. Yet a citizen may also exploit this knowledge to exert pressure on the official to decide in her favor, and an official may give in to that threat. Conceptually, it is unclear which risk dominates (Zamir and Engel, 2021). The choice between transparency and anonymity (or other types of non-transparency) is indeterminate, even if one considers only the effectiveness of the intervention.

Consequently, the policy choice requires empirical evidence. Generating such evidence is the purpose of the present paper. We translated the key elements of the policy problem into a sequential game. We implemented this game as an experiment, with a sizable sample. We have a clear result: the risk that a public official embezzles the tax proceeds is the first-order problem. Transparency is more effective than anonymity: the incidence of embezzlement is reduced from 66% to 33%. Even if transparency makes this possible, no more than 17% of all citizens retaliate against an official who has taken action over their tax evasion: in 48.9% of all cases (versus only 26.1% in the *Anonymity* condition), the normative goal is reached, and the tax is collected. This number is comfortably close to a result from a randomized control trial, where the compliance rate was found to be 44% for income not subject to a reporting requirement (Slemrod, Collins et al. 2017, 1).[14]

Based on random assignment to treatment, experiments generate reliable causal evidence. Following the tradition of experimental economics, we incentivized choices (and were completely transparent about the experiment's design). This makes the evidence even more credible. But we pay the usual price. We argue that the sequential game captures the essence of the policy conflict. But of course, the game is merely an analogue of the real-life problem that prompted the investigation. One question of particular relevance for policymakers is that we randomly assigned ordinary citizens to the role of the official. As is standard in experiments, we guaranteed anonymity between the experimenter and the participants, and among participants. The experiment captures the incentive structure of the policy problem—but in real life, the risk of corruption among public officials may be less acute, for reasons that lie beyond (ad-hoc) incentives.

In general, people's attitudes may correspond with their professional role—either because of self-selection for that role, or due to a process of socialization, or both (Heinz et al., 2005, 180, 191–194). Specifically, the *Public Service Motivation* theory suggests that self-selection in the process of joining public service results in organizations whose typical staff are disproportionately driven by prosocial concerns (Perry, 1996, Georgellis et al., 2011). The civil service might succeed in attracting particularly conscientious, or particularly risk-averse, individuals to its ranks. Moreover, long-term incentives may differ from short-term ones: the less corruption is prevalent, or visible, in a given country, the more the mere suspicion of corruption may deprive an official of career opportunities. In such a polity, corruption will be less likely in the first place.

Outside the laboratory, scrutiny by the general public is not the only measure available against inappropriate behavior or corruption by officials—internal audit and peer pressure play important roles, as well. The more effective these mechanisms, the less scrutiny by the public is critical. However the more public servants expect a harsh reaction by their supervisors when apprehended, the more they have an incentive to conceal the violation of their public duties. Typically supervisors cannot do more than random audits. In comparison, the individual citizen whose case is mishandled may not only have an informational advantage (she observes the abuse of power). She may also feel a stronger urge to go after the official whom she has observed in the act.

In our experiment, the only agent who may keep a corrupt official in check is the citizen with whom she interacts. In the field, public officials have supervisors, and the abuse of power may be picked up by the media, or by some non-governmental organization. Possibly citizens who observe an abuse of power refrain from intervening as they expect that others will audit the official anyways. If true, they would construct intervention as a contribution to a second-order public good (Yamagishi, 1986, Heckathorn, 1989), and refrain from contributing to it. While not inconceivable, we note a characteristic asymmetry: in the typical situation, the citizen whose case is mishandled has first-hand information, which gives her an informational advantage over other interventions. Moreover since it is her case, the citizen has a stronger reason to be

[14] We are grateful to an anonymous referee for drawing this parallel.

angry, and feel the urge to intervene.

Moreover, the graphical representation used in the experiment highlighted the profits and losses to the official and to the citizen—but not the harm to society at large from the loss of the tax proceeds (either because it was not paid, or because it was embezzled). This may explain the high incidence of tax evasion and embezzlement in the experiment (which, one hopes, is lower in the field). Based on our findings, we cannot say whether this impression from the field results from moral compunctions, cultural norms, or from an even greater fear of audit and penalties. At any rate, the experimental findings do suggest that transparency plays an important role in this context. It should also be noted that in experiments that are comparable in other respects, the amount of taxes paid was substantial—even when the purpose of collecting taxes was similarly undefined, and even when there was no penalty for tax evasion at all (Engel, Mittone et al. 2020).

In the *Transparency* condition, the social risk results from the fact that the citizen may inflict harm on a public official who has dutifully punished her for evading taxes. In some real-life applications, a citizen may already exert pressure before the official decides. While this alternative temporal sequence is plausible if the citizen wants to offer a side-benefit (read: a bribe), it is less plausible if the citizen wants to use a negative incentive. From the perspective of the citizen, punishing the official before she decides is pointless. The citizen has no intrinsic benefit from inflicting harm on the official. In game-theoretic lingo, punishing the official only helps the citizen if this is an outcome off the equilibrium path: the mere threat with punishment "disciplines" the official, in the sense of refraining from intervention; of course mere emotional reactions remain possible.

From an incentive perspective, the key issue is credibility. After prosecution for tax evasion has been started, executing the threat is no longer instrumental (unless the interaction is repeated, so that inflicting some harm today signals that the citizen is to be taken seriously, and might inflict even more harm in the future). In the earlier experiment by one of us on corruption, this credibility problem was overcome even in a one-shot setting. In that experiment, a citizen could offer a bribe to a public official. The official was able to bend the law in her favor. But both faced audit by an authority. If, upon audit, it was found out that the law was violated, the private benefit for the citizen was confiscated and both were fined. In addition, the citizen was able to report to the authorities, in which case both were fined with certainty. A sizeable fraction of citizens did report if the official had cashed in the bribe, but had not granted the favor. Belief data showed that experimental officials had anticipated this reaction. Critically for the present context: reporting to the authorities not only reduced the payout to the official, but also to the citizen. It was a form of costly punishment, possible only after the fact, and thereby not incentive compatible. If there was punishment regardless, it must have been motivated by behavioral effects, arguably by anger (Engel et al., 2016). It is worth noting that, in the present experiment, comparable ex post punishment (taking revenge) was much less frequently observed.

In the experiment, dyads interact. This mechanically limits the benefit of transparency to intervention by the one citizen whose taxes might be embezzled. In the field, once the identity of the public official is revealed, third parties may take notice. These third parties might even organize themselves, as in citizen watch movements. Yet as we already find a clear transparency effect in dyads, this feature of the design could at most have led us to underestimate the welfare enhancing effect of transparency.

In the experiment, interaction is one shot. In the field, taxes are due continuously. In the interest of experimental control, we have refrained from implementing a repeated game. In a repeated game, choices in later periods are influenced by experiences in earlier periods. On the cognitive side, there is room for learning (see prominently Camerer, Hua Ho, 1999). On the motivational side, there is room for reputation (Cooper et al., 1996). All of these effects could be anticipated, which is why home-grown beliefs matter, as does their updating in the light of feedback from earlier periods (see only Fischbacher and Gächter, 2010). If participants are sufficiently patient and there is uncertainty about the composition of the type space, by the folk theorem, there are multiple equilibria (Kreps, Robert, 1982). By implementing a one-shot game, we take these competing explanations out of the equation. A further advantage of the one-shot game results from the fact that we do not need to give participants feedback on the main experiment before eliciting beliefs. Consequently, beliefs cannot be contaminated by experiences in the main experiment. Finally, as Fig. 3 shows, beliefs and choices are not highly correlated. This suggests that participants did not need experiences to make up their minds regarding taking action.[15]

Yet in reality, repetition is of course conceivable. Citizens might be afraid of more intense future scrutiny after having been caught once for tax evasion. In anticipation, public officials might choose to be tough on tax evasion right from the start, as an investment into the future. On the other hand, citizens might be more willing to retaliate, to deter known officials from future intervention. Related to this, citizens and public officials might engage in tacit or explicit negotiations, using the technical power to intervene as a bargaining chip. The willingness of citizens to take revenge might be more pronounced if they have repeatedly been fined by the public official. Critically for the present purpose, the expected behavioral effects of repetition are not unidirectional. It again depends on the strength of the competing effects whether transparency or anonymity serves the public interest in collecting taxes more. It would be a worthwhile exercise to test such extensions in future work.

Another concern may be raised about the costs of action for the two participants. In the case of the official, in the experiment she bore a cost of 2 when taking action against the citizen's tax evasion. In the real world, tax officials arguably do not personally bear out of pocket costs when performing their job. The main reason for this design choice is of course experimental: monetary cost is a straightforward, easily to understand and credible, experimental manipulation. In this respect too, the experiment is only analogous to the real life. But arguably the analogy is not far-fetched. Investigating a case of potential tax evasion means more work than simply taking the tax return on file. As tax evasion is a crime, citizens accused of evasion are not unlikely to try obfuscating the evidence, which would increase the burden on the official. They might even try to fight with the official. Moreover, compared to the amount she stands to gain (10), the cost in the experiment was deliberately made low—to reflect that the official has much more to gain from embezzlement than she has to lose from doing her duty.

In the experiment, the citizen's intervention against a corrupt public official was also costly. On the one hand, the cost was not huge: the citizen lost 2 out of 10, and earned 8 rather than 10 from the experiment. In some real-life contexts, reporting to the authorities may also not be too costly. On the other hand, 2 may be perceived to be a considerable amount for an online experiment—but then again, the unpleasantness involved in filing a complaint and confronting a public official may also be quite substantial in the real world. The more a citizen has reason to suspect that many public officials are corrupt—and that those officials are happy to support each other when threatened with investigation—the higher the expected cost for a citizen taking action against them, and the more she may be deterred from reporting them. Keeping the cost low in relative terms is also appropriate since intervention does not only require effort and the willingness to accept the unpleasant experience of confronting illegal action. Intervention may also yield a concomitant benefit, like strengthening the reputation of being a vigilant citizen.

Similar concerns may be raised about the cost of retaliation by the citizen. On the one hand, in the experiment retaliation was fairly cheap: only 2. On the other hand, 2 out of the citizen's remaining endowment of 5 (after being fined 10 for her tax evasion) is a comparatively large

[15] We are grateful to an anonymous referee for pointing this out.

proportion. In some countries, this may actually reflect the expected reaction of government against attempts to threaten officials for doing their duty. The more severe the expected reaction from government, the less this potential cost of transparency may matter in the first place.

In the interest of making sure that participants understand the inevitably rich sequential game, we have framed the experiment. Framing also partly helps with external validity. Participants do not have to guess which conflict of life the experiment is meant to capture. But in the experiment, social harm is of course confined to money that does not go back to the experimenter (but is cashed in by one of the participants). In a future experiment, one might make this effect stronger by inviting passive participants who suffer if taxes are withheld or embezzled.

Finally, when summarizing the literature on experimental tests for both tax evasion and corruption, we have reported findings that differ across locations and cultures. In our experiment, we have tested a sample of participants that is rather well-educated, and predominantly lives in the UK. It could of course be that results had been different, had we tested less educated, poorer participants from a less developed economy. For such contexts, corruption is certainly an even bigger concern,[16] as is tax evasion.[17] In that respect, our results are likely only the lower bound of effects to be expected in countries with an even bigger exposure to tax evasion and corruption.

In the experiment, if the citizen cannot observe the public official, she cannot take action. In real life, merely knowing the identity of the public official may not be sufficient. This, however, seems at most a minor limitation. In nearly every legal order it is a crime if a public official pockets in money that she is supposed to spend for public projects. If the citizen can report the act and the name to prosecution, the public official is likely to be punished (if not to lose her job).

Since parameters are critical, policy makers will ultimately have to assess the local social and institutional circumstances. But unless they have reason to be fairly optimistic about public officials being law-abiding—or fairly pessimistic about citizens trying to extort a favor by threatening the individual official—our experiment cautiously suggests that Justice Brandeis was right: Sunlight is the best disinfectant.

Data Availability

We will gladly post the data once the paper is accepted

Appendix A

Model predictions

In the following, we present a simple game theoretic model. We use the model to define the conditions in which anonymity is individually beneficial (either the citizen or the public official gains higher utility), socially beneficial (with anonymity, neither is in violation of the rules) or socially detrimental (rules are observed if the identity of the public official is transparent, but are violated if that identity is concealed).

For ease of exposition, we first make the simplifying assumption that preferences are common knowledge. This provides us with a benchmark. In the next step we relax that assumption, and allow preferences to remain private information. This forces the other individual in each case to work with beliefs. As there is no objective information about the distribution of preferences, this relaxation of the assumptions introduces ambiguity. While the state space is exogenously defined (and always binary), probabilities are unknown. We theoretically define cutoffs. Beyond the respective cutoff, in the presence of the behavioral effect, predictions change.

We model the situation as the sequential game described in Fig. 1. Citizen A (red/solid) chooses between accurately reporting her income, or cheating by concealing it. If she has accurately reported, she has gross income i minus tax t. In the experiment, we set $i = 15$, $t = 5$, so that she earns 10 when correctly declaring her income, and 15 when withholding (and not reporting) income. The official B has fixed income f, and can embezzle what she has collected from A—i.e., t if A has declared income accurately. In the experiment, we set $f = 10$, so that B either earns 10 or 15. If A has been cheating, B may report her to the authorities. If she does, penalty s is meted out on A. $s > t$ makes sure that the penalty deters. In the experiment, we set $s = 10$, so A's income is reduced to $i - s = 5$. However, reporting comes at a small cost (c), so B's income is reduced to $f - c$. In the experiment we set $c = 2$, so when she reports, B only earns 8. A has the option to retaliate when reported to the authorities, at cost r to her. In the experiment, we set $r = 2$, so that her income is $i - s - r = 3$. Yet B's income is further reduced by a loss l, so that she only earns $f - c - l$. In the experiment, we set $l = 5$, so that she only earns 3. Finally, if B embezzles t, A may report to the authorities, also at cost c. Hence A earns $i - t - c$ (within the experiment's parameters of the experiment, this equals 8). Yet, the authorities impose penalties on B, so her earnings are reduced to $f + t - s$ (within the experiment's parameters, this equals 5).

If A and B have standard preferences, earnings directly translate into utility. The game is dominance solvable by backwards induction. We begin with the upper arm: $i - t - c < i - t$. If A maximizes profit, she will not report. $f + t > f$: If given the opportunity, B keeps the tax for herself. We now turn to the lower arm: $i - s - r < i - s$. A will not retaliate. $f - c < f$: B will not report. $i - t < i$: A will withhold taxes. Hence in the unique equilibrium, A cheats, and B does not report.

For the behavioral reasons discussed in the body of the paper, actual choices may differ. These behavioral effects can be captured by utility that differs from payoff. To this end, we consider three behavioral effects. First either individual may be honest. Formally, we capture this by the mental cost m. For A to be honest, we need $i - m < i - t$, or $m > t$. For B to be honest, we need $f + t - m < f$, or again $m > t$. Second, an individual may become angry, and suffer disutility a, if the other individual breaks the pertinent rule. For anger to induce the individual to report, we need $a > c$. Finally, A may want to retaliate for being reported. For this to matter, we need the positive utility v of retaliating to be greater than the pecuniary cost r.

In the first step, we assume that preferences are common knowledge: the respective other individual knows about the existence and strength of the behavioral effect. This simplifies the exposition. In the next step, we relax this (obviously counterfactual) assumption. With this assumption, the structure of the game is still as in Fig. 1. But the game is now represented in utils. Some utils change.

Table 3 summarizes the model predictions. In 6 of 12 possible situations (utility functions), the model prediction does not change in relation to standard preferences: A cheats, and B does not report. In 2 situations, A pays taxes, but B embezzles them. There are only 4 situations in which the socially desirable outcome obtains: A pays taxes, and B transfers them to the democratically chosen recipient. This is trivially the case, if A and B are honest (situation 12); if A and B are willing to report, and A does not retaliate (situation 4); if B is honest and willing to report, and A does not retaliate (situation 6); if A is honest and willing to report (situation 10).

Anonymity simultaneously closes two channels. B is not identified—so A may not report B's embezzlement, and A may not retaliate. As Table 3 shows, in 7 of 12 situations, this institutional intervention is immaterial, as the outcome is the same. Yet in 5 situations, the outcome does change. In two situations (1 and 2), one bad outcome is replaced by another: A pays her taxes, but B embezzles them. In one situation (3), the socially desirable outcome now obtains: A can no longer threaten B with retaliation. Yet in 2 situations (4 and 10), the socially desirable outcome no longer obtains: a selfish official is no longer kept in check by a citizen willing to report an abuse of public authority.

[16] https://www.transparency.org/en/cpi/2022.

[17] https://www.taxobservatory.eu/publication/global-tax-evasion-report-2024/.

Table 3
Model Predictions Assuming Common Knowledge of Preferences.

count	preferences					outcome				
	A self	B rep	A ret	B self	A rep	common knowledge				
						transp	anon	shift	pos	neg
1	1	1	1	1	1	15,10	10,15	1	0	0
2	1	1	1	1	0	15,10	10,15	1	0	0
3	1	1	1	0	0	15,10	10,10	1	1	0
4	1	1	0	1	1	10,10	10,15	1	0	1
5	1	1	0	1	0	10,15	10,15	0	0	0
6	1	1	0	0	0	10,10	10,10	0	0	0
7	1	0	0	1	1	15,10	15,10	0	0	0
8	1	0	0	1	0	15,10	15,10	0	0	0
9	1	0	0	0	0	15,10	15,10	0	0	0
10	0	0	0	1	1	10,10	10,15	1	0	1
11	0	0	0	1	0	10,15	10,15	0	0	0
12	0	0	0	0	0	10,10	10,10	0	0	0

In the next modeling step, we allow for ambiguity. In the context of the game, this ambiguity can be represented by moves of *Nature* (Fig. 5, dotted lines). Nature, with probability p_A, allows A to be a potential reporter; with probability p_B, it allows B to be a potential reporter; and with probability p_R, it allows A to retaliate.

Uncertainty about the willingness of A to retaliate is immaterial (a) if B is not willing to report or (b) A is not willing to cheat. Hence p_R only matters in situations 1–3 of Table 4. Uncertainty about B's willingness to report that A is cheating is immaterial if B is unwilling to report in the first place. Hence p_B only matters in situations 1–6 of Table 4. Uncertainty about A's willingness to report that B is embezzling the money is immaterial if A is unwilling to report it in the first place. Hence p_A only matters in situations 1, 4, 7, 10 of Table 4.

Whether or not the uncertainty affects choices depends on the utility that the individual expects to gain based on her choice, and on the probability of the respective outcome. Presumably, the individual has a fairly good estimate of her own utility, as she can forecast how she will feel if one of two possible outcomes obtains. Estimating the utility of others is much more difficult, however—but the problem is made easier by the fact that the action space of each individual, at each node of the game, is exogenously constrained to be binary. This is why the counterpart does not need a cardinal estimate of the utility of her counterpart. A, for example, does not need to know u_B (A cheats; B remains passive) or u_B (A cheats, B reports) – u_B (A cheats, B remains passive). All she needs to know is p_B—i.e., the probability that this difference is larger than the cost c of reporting. It is not important for A whether this difference is just slightly larger than c, or much larger. Similarly, B only needs to estimate p_A and—if she is willing to report—of p_R. Hence B does not need to estimate A's utility from reporting, or from retaliating.

There is one final complication. With respect to any of these risks, A or B may not be neutral. If they are risk-averse, they would put more weight on the outcome that is bad, rather than good, for them. B may be even more concerned about retaliation, or of being reported to the authorities. A would be particularly concerned about being penalized. Empirically, disentangling the cognitive effect (the estimate of the respective probability) and the motivational effect (attitude toward the risk in question) may be challenging. Yet for the purposes of this project, we can contain this complication, as we are ultimately only interested in the effects on choices. This is why we can represent risk aversion by a higher subjective estimate of the probability that the bad outcome obtains.

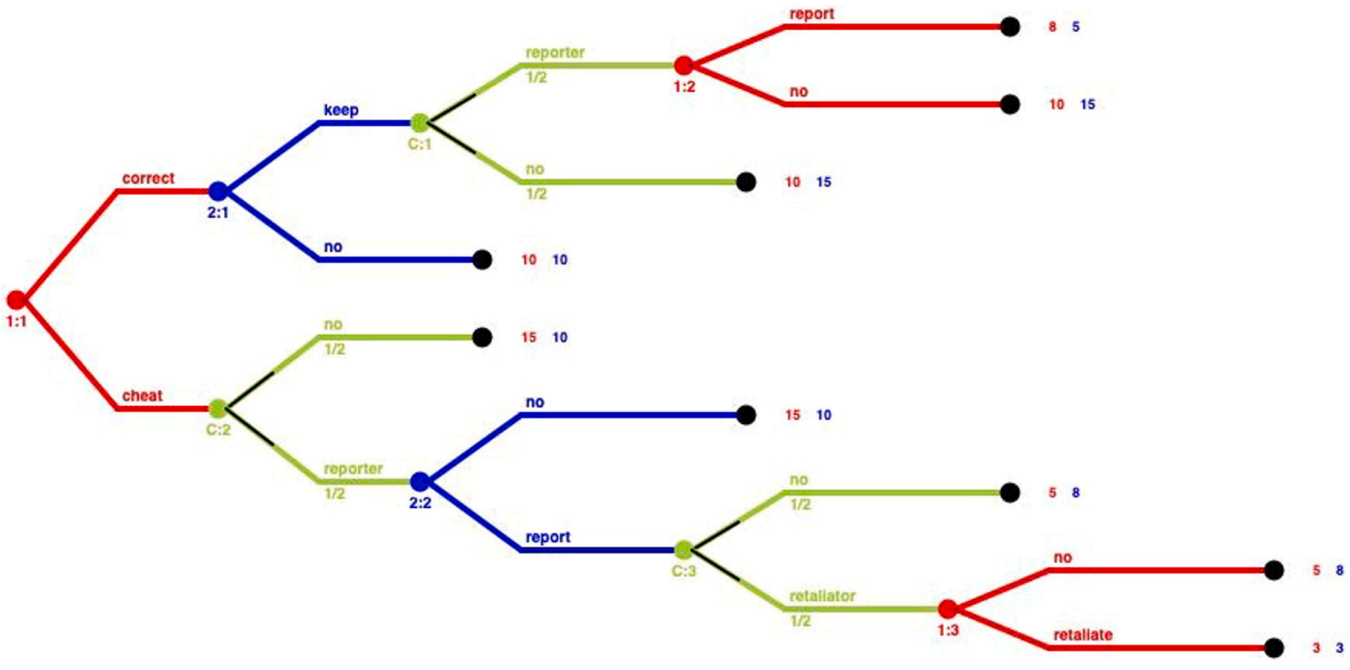

Fig. 5. Game Tree Allowing for Preference Uncertainty.

Table 4
Model Predictions Assuming Preference Uncertainty.

case	preferences					outcome in the face of uncertainty					
						transparency			anonymity		
	A self	B rep	A ret	B self	A rep	15,10	10,15	10,10	15,10	10,15	10,10
1	1	1	1	1	1	$p_B\|p_R$ <.5	$p_B\|p_R$ >.5 p_A<.5	$p_B\|p_R$ >.5 p_A>.5	p_B<.5	p_B>.5	0
2	1	1	1	1	0	$p_B\|p_R$ <.5	$p_B\|p_R$ >.5	0	p_B<.5	p_B>.5	0
3	1	1	1	0	0	$p_B\|p_R$ <.5	0	$p_B\|p_R$ >.5	p_B<.5	0	p_B>.5
4	1	1	0	1	1	p_B<.5 p_A>.5	p_B>.5 p_A<.5	p_B>.5 p_A>.5	p_B<.5	p_B>.5	0
5	1	1	0	1	0	p_B<.5	p_B>.5	0	p_B<.5	p_B>.5	0
6	1	1	0	0	0	p_B<.5	0	p_B>.5	p_B<.5	0	p_B>.5
7	1	0	0	1	1	1	0	0	1	0	0
8	1	0	0	1	0	1	0	0	1	0	0
9	1	0	0	0	0	1	0	0	1	0	0
10	0	0	0	1	1	0	p_A<.5	p_A>.5	0	1	0
11	0	0	0	1	0	0	1	0	0	1	0
12	0	0	0	0	0	0	0	1	0	0	1

Despite the uncertainty, a selfish B will not embezzle, if $p_A(f + t - s) + (1 - p_A)(f + t) < f$, or if the estimated probability that A will report to the authorities is larger than t/s. All three parameters (f, t, s) are induced by the experiment's design. With these parameters, B will behave honestly, regardless of the temptation, if her assessment of A reporting her embezzlement is greater than ½. By the same logic, if A is not prepared to retaliate, she will pay her taxes if $p_B(i - s) + (1 - p_B)i < i - t$, or if the estimated probability that B would report tax evasion is larger than t/s. Again, all parameters (i, t, s) are induced by the experiment's design: with these parameters, A will behave honestly, regardless of the temptation, if her assessment of B reporting her tax evasion is larger than ½.

If B expects the probability p_R that A would retaliate if she reports her tax evasion to the authorities to be positive, she needs one parameter that should be easy for her to estimate, but which the experimenter cannot directly observe—namely, her gain in utility from preventing A from getting away with it, which we denote as x. Despite this uncertainty, B reports tax evasion, provided $p_R(f - c + x - l) + (1 - p_R)(f - c + x) > f$, or if the estimated probability p_R that A will retaliate is smaller than $(x - c)/l$. If x is small (say, 4), B would only report if $p_R < 2/5$. If x is large (say, 6), B would report if $p_R < 4/5$. If x is very large (7 or larger), B would always report, despite the risk of retaliation.

For B's choices, the estimates of p_A and p_R suffice. However, when A decides whether or not to cheat, she must estimate $p_B|p_R$. Now p_R is B's subjective estimate of A retaliating. A thus needs an *estimate of this estimate*—or a second-order belief. However, in fact, this complication is minor, as p_B is a subjective probability in the first place. If A is selfish herself, for the purposes of her decision, only $p_B > ½$ matters—whether p_B is unconditional, or conditional on A's estimate of B's estimate of retaliation.

Table 4 brings together model predictions based on subjective probability estimates, and on *Transparency* vs. *Anonymity*. When compared with Table 3, it demonstrates that uncertainty can be beneficial: if preferences are common knowledge, the socially desirable outcome (A pays taxes, B transfers the proceeds to the intended recipient) can be reached in 4 situations (4, 6, 10, 12). With preference uncertainty, this is possible in 6 situations (1, 3, 4, 6, 10, 12). Nonetheless, the socially beneficial outcome becomes more precarious. Apart from when A and B are honest in the first place (situation 12), it is only reached when the estimated probability of being reported by the other individual is >.5.

There is still one situation in which anonymity is socially beneficial, in the sense of reaching the socially desirable outcome (situation 3). Yet with preference uncertainty, the beneficial effect is less pronounced. Rather than the (arguably more precarious) conditional probability $p_B|p_R$, only the unconditional probability p_B must be greater than ½. As when preferences are common knowledge, with preference uncertainty anonymity is detrimental in situations 4 and 10—the possibility that the socially desirable outcome is reached disappears. Yet in situation 4, it depends on p_B—regardless of whether this is to the benefit of A or B. If there were otherwise room for retaliation, anonymity gives B a comparative advantage. If $p_B|p_R$ <.5, but p_B>.5, her selfishness determines the outcome.

The purpose of this project was to test the effect of anonymity. The above analysis demonstrates that the effect depends on which normative problem looms largest: if $m_B > m_A$, it is more important to keep the citizen in check, rather than the official. The world is more likely to be in situation 3, 6 or 9. We may also have $a_B > a_A$: the public official is more likely to bear the cost of policing A, than vice versa. The world is more likely to be in situations 2, 3, 5 or 6 than in situations 4, 7 or 10. Finally, the official may be strongly concerned about her estimate of v—namely, about the risk of retaliation. This would matter if the world is in situations 1 – 3.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.irle.2024.106189.

References

Abbink, Klaus, Irlenbusch, Bernd, Renner, Elke, 2002. An experimental bribery game. J. Law, Econ. Organ. 18, 428–454.

Abbink, Klaus, Serra, Danila, 2012. Anticorruption policies. lessons from the lab. Res. Exp. Econ. 15, 77–115.

Abbink, Klaus, Wu, Kevin, 2017. Reward self-reporting to deter corruption. An experiment on mitigating collusive bribery. J. Econ. Behav. Organ. 133, 256–272.

Abraham, Martin, Lorek, Kerstin, Richter, Friedemann, Wrede, Matthias, 2017. Collusive tax evasion and social norms. Int. Tax. Public Financ. 24 (2), 179–197.

Ackert, Lucy, F., Martinez-Vazquez, Jorge, Rider, Mark, 2007. Social preferences and tax policy design: some experimental evidence. Econ. Inq. 45 (3), 487–501.

Alatas, Vivi, Cameron, Lisa, Chaudhuri, Ananish, Erkal, Nisvan, Gangadharan, Lata, 2009a. Gender, culture, and corruption. Insights from an experimental analysis. South. Econ. J. 75, 663–680.

Alatas, Vivi, Cameron, Lisa, Chaudhuri, Ananish, Erkal, Nisvan, Gangadharan, Lata, 2009b. Subject pool effects in a corruption experiment. A comparison of Indonesian public servants and Indonesian students. Exp. Econ. 12 (1), 113–132.

Allingham, Michael G., Sandmo, Agnar, 1972. Income tax evasion. A theoretical analysis. J. Public Econ. 1 (3–4), 323–338.

Alm, James, 1991. A perspective on the experimental analysis of taxpayer reporting. Account. Rev. 66 (3), 577–593.

Alm, James, 2012. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. Int. Tax. Public Financ. 19 (1), 54–77.

Alm, James, Jackson, Betty, McKee, Michael, 1992a. Institutional uncertainty and taxpayer compliance. Am. Econ. Rev. 82 (4), 1018–1026.

Alm, James, Jackson, Betty R., McKee, Michael, 1992b. Estimating the determinants of taxpayer compliance with experimental data. Natl. Tax. J. 45, 107–114.
Alm, James, Sanchez, Isabel, De Juan, Ana, 1995. Economic and noneconomic factors in tax compliance. Kyklos 48 (1), 1–18.
Almenberg, Johan, Dreber, Anna, Apicella, Coren L., Rand, David, 2011. Third party reward and punishment. group size, efficiency and public goods. Psychology of punishment. N. Palmetti J. P. Russo. N. Y., Nova 73–92.
Anderhub, Vital, Giese, Sebastian, Güth, Werner, Hoffmann, Antje, Otto, Thomas, 2001. Tax evasion with earned income-an experimental study. Publ. Finance Anal. 58, 188–206.
Andreoni, James, Erard, Brian, Feinstein, Jonathan, 1998. Tax compliance. J. Econ. Lit. 36, 818–860.
Ariel, Barak, 2012. Deterrence and moral persuasion effects on corporate tax compliance. Findings from a randomized controlled trial. Criminology 50 (1), 27–69.
Armantier, Olivier, Boly, Amadou, 2011. A controlled field experiment on corruption. Eur. Econ. Rev. 55 (8), 1072–1082.
Armantier, Olivier, Boly, Amadou, 2013. Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. Econ. J. 123 (573), 1168–1187.
Armantier, Olivier, Boly, Amadou, Serra, D., Wantchekon, L., 2012. On the external validity of laboratory experiments on corruption. N. Adv. Exp. Res. Corrupt. D. Serra L. Wantchekon, Emerald 117–144.
Balafoutas, Loukas, 2011. Public beliefs and corruption in a repeated psychological game. J. Econ. Behav. Organ. 78 (1), 51–59.
Balafoutas, Loukas, Beck, Adrian, Kerschbamer, Rudolf, Sutter, Matthias, 2015. The hidden costs of tax evasion. Collaborative tax evasion in markets for expert services. J. Public Econ. 129, 14–25.
Balafoutas, Loukas, Grechenig, Kristoffel, Nikiforakis, Nikos, 2014. Third-party punishment and counter-punishment in one-shot interactions. Econ. Lett. 122 (2), 308–310.
Baldry, Jonathan C., 1986. Tax evasion is not a gamble. A report on two experiments. Econ. Lett. 22 (4), 333–335.
Banerjee, Ritwik, 2016. On the interpretation of bribery in a laboratory corruption game. Moral frames and social norms. Exp. Econ. 19, 240–267.
Barr, Abigail, Lindelow, Magnus, Serneels, Pieter, 2009. Corruption in public service delivery. An experimental analysis. J. Econ. Behav. Organ. 72 (1), 225–239.
Barr, Abigail, Serra, Danila, 2009. The effects of externalities and framing on bribery in a petty corruption experiment. Exp. Econ. 12 (4), 488–503.
Barr, Abigail, Serra, Danila, 2010. Corruption and culture. An experimental analysis. J. Public Econ. 94 (11), 862–869.
Batrancea, Larissa, Nichita, Anca, Olsen, Jerome, Kogler, Christoph, Kirchler, Erich, Hoelzl, Erik, Weiss, Avi, Torgler, Benno, Fooken, Jonas, Fuller, Joanne, 2019. Trust and power as determinants of tax compliance across 44 nations. J. Econ. Psychol. 74, 102191.
Battiston, Pietro, Gamba, Simona, 2013. Is tax compliance a social norm? A field experiment. Univ. Milan. Bicocca Dep. Econ., Manag. Stat. Work. Pap.
Bayer, Ralph-C., Sutter, Matthias, 2009. The excess burden of tax evasion. An experimental detection–concealment contest. Eur. Econ. Rev. 53 (5), 527–543.
Bazart, Cécile, Bonein, Aurélie, 2014. Reciprocal relationships in tax compliance decisions. J. Econ. Psychol. 40, 83–102.
Beekman, Gonne, Bulte, Erwin, Nillesen, Eleonora, 2014. Corruption, investments and contributions to public goods. Experimental evidence from rural liberia. J. Public Econ. 115, 37–47.
Bilotkach, Volodymyr (2005). A Tax Evasion-Bribery Game. Experimental Evidence from Ukraine.
Blackwell, Calvin (2007). A Meta-Analysis of Tax Compliance Experiments.
Blaufus, Kay, Hundsdoerfer, Jochen, Jacob, Martin, Sünwoldt, Matthias, 2016. Does legality matter? The case of tax avoidance and evasion. J. Econ. Behav. Organ. 127, 182–206.
Blumenthal, Marsha, Christian, Charles, Slemrod, Joel, Smith, Matthew G., 2001. Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. Natl. Tax. J. 54, 125–138.
Blundell, Wesley, 2020. When threats become credible. A natural experiment of environmental enforcement from Florida. J. Environ. Econ. Manag. 101, 102288.
Bobek, Donna D., Amy, M.Hageman, Kelliher, Charles F., 2013. Analyzing the role of social norms in tax compliance behavior. J. Bus. Ethics 115, 451–468.
Bobek, Donna D., Hatfield, Richard C., 2003. An investigation of the theory of planned behavior and the role of moral obligation in tax compliance. Behav. Res. Account. 15 (1), 13–38.
Bobek, Donna D., Robin, W.Roberts, Sweeney, John T., 2007. The social norms of tax compliance: evidence from australia, singapore, and the United States. J. Bus. Ethics 74, 49–64.
Bobkova, Nina, Egbert, Henrik, 2013. Corruption investigated in the lab. A survey of the experimental literature. Int. J. Latest Trends Financ. Econ. Sci. 2 (4), 337–349.
Bordignon, Massimo, 1993. A Fairness approach to income tax evasion. J. Public Econ. 52 (3), 345–362.
Bosco, Luigi, Mittone, Luigi, 1997. Tax evasion and moral constraints. Some experimental evidence. Kyklos 50 (3), 297–324.
Bott, Maria, Kristina, Cappelen, Alexander W., Sorensen, Erik, Tungodden, Bertil, 2020. You've got mail. A randomised field experiment on tax evasion. Manag. Sci. 66, 2801–2819.
Bovens, Marc (2005). Public Accountability. Oxford Handbook of Public Management. E. Ferlie, L. E. Lynn Jr and C. Pollitt, Oxford University Press: 182-208.
Brandeis, Louis D. (1914). Other People's Money and How the Bankers Use It, Cosimo, Inc.
Cadsby, C.Bram, Maynes, Elizabeth, Trivedi, Viswanath Umashanker, 2006. Tax compliance and obedience to authority at home and in the lab. A new experimental approach. Exp. Econ. 9 (4), 343–359.
Callen, Michael, Long, James D., 2015. Institutional corruption and election fraud. Evidence from a field experiment in Afghanistan. Am. Econ. Rev. 105 (1), 354–381.
Camerer, Colin F., Hua Ho, Teck, 1999. Experience-weighted attraction learning in normal form games. Econometrica 67 (4), 827–874.
Cameron, Lisa A., Ananish Chaudhuri, Nisvan Erkal and Lata Gangadharan (2005). Do Attitudes Towards Corruption Differ across Cultures? Experimental Evidence from Australia, India, Indonesia and Singapore.
Campanario, Juan Miguel, 1998a. Peer review for journals as it stands today—part 1. Sci. Commun. 19 (3), 181–211.
Campanario, Juan Miguel, 1998b. Peer review for journals as it stands today—part 2. Sci. Commun. 19 (4), 277–306.
Casagrande, Alberto, Di Cagno, Daniela, Pandimiglio, Alessandro, Spallone, Marco, 2015. The effect of competition on tax compliance. The role of audit rules and shame. J. Behav. Exp. Econ. 59, 96–110.
Casal, Sandro, Faillo, Marco, Mittone, Luigi, 2022. I want to pay!-Identifying the Unconditional Tax Propensity (Utp). J. Econ. Behav. Organ. 197, 103–114.
Cason, Timothy, N., Gangadharan, Lata, 2006. An experimental study of compliance and leverage in auditing and regulatory enforcement. Econ. Inq. 44 (2), 352–366.
Castro, Lucio, Scartascini, Carlos, 2015. Tax compliance and enforcement in the pampas. Evidence from a field experiment. J. Econ. Behav. Organ. 116, 65–82.
Chen, Daniel L., Schonger, Martin, Wickens, Chris, 2016. Otree—an open-source platform for laboratory, online, and field experiments. J. Behav. Exp. Financ. 9, 88–97.
Chetty, Raj, Looney, Adam, Kroft, Kory, 2009. Salience and taxation. Theory and evidence. Am. Econ. Rev. 99 (4), 1145–1177.
Chong, Alberto, De La O, Ana L., Karlan, Dean, Wantchekon, Leonard, 2015. Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification. J. Polit. 77 (1), 55–71.
Christian, Roberta Calvet, Alm, James, 2014. Empathy, sympathy, and tax compliance. J. Econ. Psychol. 40, 62–82.
Coleman, Stephen (2007). The Minnesota Income Tax Compliance Experiment. Replication of the Social Norms Experiment.
Cooper, Russell, Douglas, V.De.Jong, Forsythe, Robert, Ross, Thomas, 1996. Cooperation without reputation. Experimental evidence from prisoner's dilemma games. Games Econ. Behav. 12, 187–218.
Corbacho, Ana, Daniel W.Gingerich, Virginia Oliveros, Mauricio Ruiz-Vega, 2016. Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica. Am. J. Political Sci. 60, 1077–1092.
Coricelli, Giorgio, Joffily, Mateus, Montmarquette, Claude, Villeval, Marie Claire, 2010. Cheating, emotions, and rationality. An experiment on tax evasion. Exp. Econ. 13 (2), 226–247.
Coricelli, Giorgio, Rusconi, Elena, Villeval, Marie Claire, 2014. Tax evasion and emotions. An empirical test of re-integrative shaming theory. J. Econ. Psychol. 40, 49–61.
Cullis, John, Jones, Philip, Soliman, Amal, 2012. Spite effects' in tax evasion experiments. J. Socio-Econ. 41 (4), 418–423.
Cummings, Ronald G., Martinez-Vazquez, Jorge, McKee, Michael, Torgler, Benno, 2009. Tax morale affects tax compliance. Evidence from surveys and an artefactual field experiment. J. Econ. Behav. Organ. 70 (3), 447–457.
Doerrenberg, Philipp, 2015. Does the use of tax revenue matter for tax compliance behavior? Econ. Lett. 128, 30–34.
Doerrenberg, Philipp, Duncan, Denvil, 2014. Distributional implications of tax evasion. Evidence from the lab. Public Financ. Rev. 42, 720–744.
Doerrenberg, Philipp and Andreas Peichl (2017). Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment.
Dulleck, Uwe, Fooken, Jonas, Newton, Cameron, Ristl, Andrea, Schaffner, Markus, Torgler, Benno, 2016. Tax compliance and psychic costs. Behavioral experimental evidence using a physiological marker. J. Public Econ. 134, 9–18.
Dwenger, Nadja, Henrik Kleven, Imran Rasul, Johannes Rincke, 2016. Extrinsic Vs Intrinsic Motivations for Tax Compliance. Evidence from a Randomized Field Experiment in Germany. Am. Econ. J. Econ. Policy 8, 203–232.
Enachescu, Janina, Olsen, Jerome, Kogler, Christoph, Zeelenberg, Marcel, Breugelmans, Seger M., Kirchler, Erich, 2019. The role of emotions in tax compliance behavior. A mixed-methods approach. J. Econ. Psychol. 74, 102194.
Engel, Christoph, 2015. Tacit collusion. The neglected experimental evidence. J. Empir. Leg. Stud. 12, 537–577.
Engel, Christoph, Sebastian, J.Goerg, Yu, Gaoneng, 2016. Symmetric vs. asymmetric punishment regimes for collusive bribery. Am. Law Econ. Rev. 18, 506–556.
Engel, Christoph, Mittone, Luigi, Morreale, Azzurra, 2020. Tax morale and fairness in conflict. An experiment. J. Econ. Psychol. 81, 102314.
Engel, Christoph, Luigi Mittone and Azzurra Morreale (2024). Outcomes or Participation? Experimentally testing competing sources of legitimacy for taxation. Economic Inquiry **(forthcoming)**.
Engel, Christoph, David, G.Rand, 2014. What does "clean" really mean? The implicit framing of decontextualized experiments. Econ. Lett. 122 (3), 386–389.
Engel, Christoph, Reuben, Alicja, 2015. The people's hired guns? Experimentally testing the motivating force of a legal frame. Int. Rev. Law Econ. 43, 67–82.
Engel, Christoph, Zhurakhovska, Lilia, 2017. You are in charge. Experimentally testing the motivating power of holding a judicial office. J. Leg. Stud. 46, 1–50.
Falkinger, Josef, 1995. Tax evasion, consumption of public goods and fairness. J. Econ. Psychol. 16 (1), 63–72.
Fehr, Ernst, Fischbacher, Urs, 2004. Third-party punishment and social norms. Evol. Hum. Behav. 25, 63–87.

Fehr, Ernst, Gächter, Simon, 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90, 980–994.
Fehr, Ernst, Gächter, Simon, 2002. Altruistic punishment in humans. Nature 415, 137–140.
Filer, Randall K., Jan Hanousek, Tomas Lichard and Karine Torosyan (2016). 'Flattening' the Tax Evasion. Evidence from the Post-Communist Natural Experiment.
Fischbacher, Urs, Gächter, Simon, 2010. Social preferences, beliefs, and the dynamics of free riding in public good experiments. Am. Econ. Rev. 100, 541–556.
Fochmann, Martin, Eike, B.Kroll, 2016. The effects of rewards on tax compliance decisions. J. Econ. Psychol. 52, 38–55.
Fochmann, Martin, Wolf, Nadja, 2019. Framing and salience effects in tax evasion decisions. An experiment on underreporting and overdeducting. J. Econ. Psychol. 72, 260–277.
Gangl, Katharina, Torgler, Benno, Kirchler, Erich, Hofmann, Eva, 2014. Effects of supervision on tax compliance. Evidence from a field experiment in Austria. Econ. Lett. 123 (3), 378–382.
Garcia, Filomena, Opromolla, Luca David, Vezzulli, Andrea, Marques, Rafael, 2020. The effects of official and unofficial information on tax compliance. J. Econ. Psychol., 102265
García-Gallego, Aurora, Georgantzis, Nikos, Jaber-López, Tarek, Michailidou, Georgia, 2020. Audience effects and other-regarding preferences against corruption: experimental evidence. J. Econ. Behav. Organ. 180, 159–173.
Gardner, James A., 2010. Anonymity and democratic citizenship. William Mary Bill. Rights J. 19, 927–958.
Georgellis, Elisabetta Iossa, Yannis, Tabvuma, Vurain, 2011. Crowding out intrinsic motivation in the public sector. J. Public Adm. Res. Theory 21 (3), 473–493.
Güth, Werner, Sausgruber, Rupert, 2008. Voting between tax regimes to fund a public good. Econ. Gov. 9 (4), 287–303.
Hanno, D.M., Violette, G.R., 1996. An analysis of moral and social influences in tax compliance. Behav. Res. Account. 8, 57–75.
Heath, Chip, Tversky, Amos, 1991. Preference and belief. ambiguity and competence in choice under uncertainty. J. Risk Uncertain. 4, 5–28.
Heckathorn, Douglas D., 1989. Collective action and the second-order free-rider problem. Ration. Soc. 1, 78–100.
Heinz, John P., Robert L.Nelson, Rebecca L. Sandefur and Edward O. Laumann (2005). Urban Lawyers. The New Social Structure of the Bar, University of Chicago Press.
Hood, Christopher, 2010. Accountability and transparency. Siamese twins, matching parts, awkward couple? West Eur. Polit. 33 (5), 989–1009.
Incerti, Trevor, 2020. Corruption Information and vote share. a meta-analysis and lessons for experimental design. Am. Political Sci. Rev. 114 (3), 761–774.
Jackson, Betty R., Jones, Sally, 1985. Salience of tax evasion penalties versus detection risk. J. Am. Tax. Assoc. 6 (2), 7–17.
Jimenez, Peggy, Iyer, Govind S., 2016. Tax compliance in a social setting. The influence of social norms, trust in government, and perceived fairness on taxpayer compliance. Adv. Account. 34, 17–26.
Kao, Chieh (2016). No Taxation with Information on Representation. A Survey Experiment on Tax Morale in China.
Khadjavi, Menusch, Lange, Andreas, Nicklisch, Andreas, 2017. How transparency may corrupt. Experimental evidence from asymmetric public goods games. J. Econ. Behav. Organ. 142, 468–481.
Kirchler, Erich, 2007. The Economic Psychology of Tax Behaviour. Cambridge University Press,.
Kleven, Jacobsen, Henrik, Knudsen, Martin B., Thustrup Kreiner, Claus, Pedersen, S.øren, Saez, Emmanuel, 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica 79 (3), 651–692.
Kotakorpi, Kaisa and Jani-Petri Laamanen (2016). Prefilled Income Tax Returns and Tax Compliance. Evidence from a Natural Experiment.
Kreps, David M., Robert, B.Wilson, 1982. Sequential equilibria. Econometrica 50, 863–894.
Lambsdorff, Johann Graf, 2012. Behavioral and experimental economics as a guidance to anticorruption. N. Adv. Exp. Res. Corrupt. Res. Exp. Econ. 15, 279–299.
Lambsdorff, Johann Graf, Frank, Björn, 2011. Corrupt reciprocity. Experimental evidence on a men's game. Int. Rev. Law Econ. 31 (2), 116–125.
Lefebvre, Mathieu, Pestieau, Pierre, Riedl, Arno, Villeval, Marie Claire, 2015. Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. Int. Tax. Public Financ. 22 (3), 401–425.
Leibbrandt, Andreas, López-Pérez, Raúl, 2012. An exploration of third and second party punishment in ten simple games. J. Econ. Behav. Organ. 84, 753–766.
Lergetporer, Philipp, Angerer, Silvia, Glätzle-Rützler, Daniela, Sutter, Matthias, 2014. Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. Proc. Natl. Acad. Sci. 111 (19), 6916–6921.
Luttmer, Erzo F.P., Singhal, Monica, 2014. Tax morale. J. Econ. Perspect. 28 (4), 149–168.
Machina, Mark J., Siniscalchi, Marciano, 2014. Ambiguity and ambiguity aversion. In: Handbook of the Economics of Risk and Uncertainty, **1**. Elsevier,, pp. 729–807.
Magessi, Trindade, Nuno, Antunes, Luis, 2015. Risk perception and risk attitude on a tax evasion context. Cent. Eur. J. Econ. Model. Econ. 7 (3), 127–149.
Mansour, Sarah, Wallace, Sally, Sadiraj, Vjollca, Hassan, Mazen, 2021. How do electoral and voice accountability affect corruption? Experimental evidence from Egypt. Eur. J. Political Econ. 68, 101994.
McCaffery, Edward (2000). Cognitive Theory and Tax. Behavioral Law and Economics. C. R. Sunstein. Cambridge, Cambridge University Press: 398-421.
McGraw, Kathleen and John Scholz, 1991. Appeals to civic virtue versus attention to self-interest: effects on tax compliance. Law Soc. Rev. 25, 471–493.
McKee, Michael and Christian A. Vossler (2013). Tax Withholding and Tax Compliance: Evidence from a Framed Field Experiment1.
Niesiobędzka, Ma.łgorzata, Kołodziej, Sabina, 2019. The impact of procedural fairness and extent of a tax loss or gain on the acceptance of tax authority decisions and the intention to appeal against them. Psychol. Public Policy, Law 25 (1), 46–56.
Nikiforakis, Nikos, 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? J. Public Econ. 92, 91–112.
Nikiforakis, Nikos, Mitchell, Helen, 2014. Mixing the carrots with the sticks. Third party punishment and reward. Exp. Econ. 17 (1), 1–23.
Olken, Benjamin A., 2007. Monitoring corruption. Evidence from a field experiment in Indonesia. J. Political Econ. 115, 200–249.
Peisakhin, Leonid, Pinto, Paul, 2010. Is Transparency an effective anti-corruption strategy? Evidence from a field experiment in India. Regul. Gov. 4 (3), 261–280.
Perez-Truglia, Ricardo, Troiano, Ugo, 2015. Shaming tax delinquents. Theory and evidence from a field experiment in the United States. National Bureau of Economic Research.
Perry, James L., 1996. Measuring public service motivation. An assessment of construct reliability and validity. J. Public Adm. Res. Theory 6 (1), 5–22.
Pickhardt, Michael, Prinz, Aloys, 2014. Behavioral dynamics of tax evasion. A survey. J. Econ. Psychol. 40, 1–19.
Rockenbach, Bettina, Milinski, Manfred, 2006. The efficient interaction of indirect reciprocity and costly punishment. Nature 444 (7120), 718–723.
Rodriguez-Justicia, David, Theilen, Bernd, 2018. Education and tax morale. J. Econ. Psychol. 64, 18–48.
Ryvkin, Dmitry, Serra, Danila, 2020. Corruption and competition among bureaucrats. An experimental study. J. Econ. Behav. Organ. 175, 439–451.
Sælen, Håkon, Kallbekken, Steffen, 2011. A choice experiment on fuel taxation and earmarking in Norway. Ecol. Econ. 70 (11), 2181–2190.
Salmon, Timothy C., Serra, Danila, 2017. Corruption, social judgment and culture: an experiment. J. Econ. Behav. Organ. 142, 64–78.
Savage, Leonard J., 1954. The foundations of statistics. Wiley, New York.
Selten, Reinhard, 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. Beiträge zur exp. Wirtschaftsforschung. E. Sauermann. Tübingen, Mohr 136–168.
Serra, Danila, 2012. Combining top-down and bottom-up accountability. Evidence from a bribery experiment. J. Law, Econ., Organ. 28 (3), 569–587.
Slemrod, Joel, Blumenthal, Marsha, Christian, Charles, 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. J. Public Econ. 79 (3), 455–483.
Slemrod, Joel, Collins, Brett, Hoopes, Jeffrey L., Reck, Daniel, Sebastiani, Michael, 2017. Does credit-card information reporting improve small-business tax compliance? J. Public Econ. 149, 1–19.
Spicer, Michael, W., Everett Thomas, J., 1982. Audit probabilities and the tax evasion decision. An experimental approach. J. Econ. Psychol. 2 (3), 241–245.
Sundström, Aksel, 2012. Corruption and regulatory compliance. Experimental findings from south african small-scale fisheries. Mar. Policy 36 (6), 1255–1264.
Swenson, Charles W., 1988. Taxpayer behavior in response to taxation. An experimental analysis. J. Account. Public Policy 7 (1), 1–28.
Tan, Fangfang, Yim, Andrew, 2014. Can strategic uncertainty help deter tax evasion? An experiment on auditing rules. J. Econ. Psychol. 40, 161–174.
Telle, Kjetil, 2013. Monitoring and enforcement of environmental regulations: lessons from a natural field experiment in Norway. J. Public Econ. 99, 24–34.
Torgler, Benno, 2003. Beyond punishment. A tax compliance experiment with taxpayers in Costa Rica. Revista di Análisis Económico 18 (1), 27–56.
Torgler, Benno, 2004. Moral Suasion. An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. Econ. Gov. 5 (3), 235–253.
Torgler, Benno, 2007. Tax Compliance and Tax Morale. A Theoretical and Empirical Analysis. Edward Elgar Publishing.
Traxler, Christian, 2010. Social norms and conditional cooperative taxpayers. Eur. J. Political Econ. 26, 89–103.
Tversky, Amos, Kahneman, Daniel, 1981. The framing of decisions and the psychology of choice. Science 211, 453–458.
Tversky, Amos, Kahneman, Daniel, 1986. Rational choice and the framing of decisions. J. Bus. 59, 5251–5278.
Tversky, Amos, Simonson, Itamar, 1993. Context-dependent preferences. Manag. Sci. 39, 117–185.
Wahl, Ingrid, Muehlbacher, Stephan, Kirchler, Erich, 2010. The impact of voting on tax payments. Kyklos 63, 144–158.
Wan, Junmin, 2010. The incentive to declare taxes and tax revenue. The lottery receipt experiment in China. Rev. Dev. Econ. 14 (3), 611–624.
Weber, Martin H., Camerer, Colin F., 1992. Recent developments in modeling preferences. Uncertainty and ambiguity. J. Risk Uncertain. 5 (4), 325–370.
Yamagishi, Toshio, 1986. The provision of a sanctioning system as a public good. J. Personal. Soc. Psychol. 51, 110–116.
Zamir, Eyal, Engel, Christoph, 2021. Sunlight is the best disinfectant-or is it? Anonymity as a means to enhance impartiality. Ariz. Law Rev. 63, 1063–1087.
Zhang, Nan, 2015. Changing a 'culture' of corruption: evidence from an economic experiment in Italy. Ration. Soc. 27 (4), 387–413.
Zhang, Nan, 2018. Institutions, norms, and accountability. A corruption experiment with Northern and Southern Italians. J. Exp. Political Sci. 5 (1), 11–25.
Zhang, Nan, Andrighetto, Giulia, Ottone, Stefania, Ponzano, Ferruccio, Steinmo, Sven, 2016. Willing to Pay? Tax compliance in Britain and Italy: an experimental analysis. PloS One 11 (2), e0150277.