

ARTICLE

Bringing tax avoiders to light: moral framing and shaming in a public goods experiment

Stefanos A. Tsikas 

Hannover Medical School, Hannover, Germany
E-mail: tsikas.stefanos@mh-hannover.de

(Received 7 August 2020; revised 27 January 2021; accepted 3 February 2021;
first published online 26 March 2021)

Abstract

With a linear public goods game played in six different variants, this article studies two channels that might moderate social dilemmas and increase cooperation without using pecuniary incentives: moral framing and shaming. We find that cooperation is increased when noncontributing to a public good is framed as morally debatable and socially harmful tax avoidance, while the mere description of a tax context has no effect. However, without social sanctions in place, cooperation quickly deteriorates due to social contagion. We find ‘shaming’ free-riders by disclosing their misdemeanor to act as a strong social sanction, irrespective of the context in which it is applied. Moralizing tax avoidance significantly reinforces shaming, compared with a simple tax context.

JEL codes: E62; H26; H30

Keywords: shaming; framing; tax avoidance; social contagion; public goods game

Introduction

During the past few years, the media has repeatedly reported on large-scale tax avoidance schemes by firms and wealthy individuals, often assisted by wealth management firms and the professional classes. Some of the reported activities may constitute criminal tax evasion, fraud, or money laundering – but most seem perfectly legal. Still, the press and the wider public judge their escape from taxes as morally reprehensible, opportunistic, and disreputable. The incriminated persons may not have violated any law, but their behavior is seen as debasing the spirit of the tax law for the sake of personal gains. The broad media coverage has not only brought that behavior to light, but it overwhelmingly also issued a devastating moral verdict. For the politicians, sports stars, artists, or business people who were pilloried for tax avoidance, the unprecedented and embarrassing publicity damaged their reputation – and the shaming might deter them and others in the future.

These developments exemplify two insights: first, socially cooperative behavior often cannot be fully formalized in laws, legislated, and judicially enforced. Rather it requires a specific ‘morality’ – a sense of virtue and decency, of duty and civic

© The Author(s), 2021. Published by Cambridge University Press

obligations – or a normatively ‘right’ demeanor. Appeals to this morality may positively affect individuals’ prosocial behavior. Second, to reinforce or induce socially warranted behavior, informal reputational mechanisms – for example, the naming and shaming of alleged tax dodgers – may play an important role. The experience or already the fear of being visibly identified to the general public or to one’s peers as noncooperative and opportunistic can make individuals act more cooperatively.

We report the results of a Public Goods Game that interacts (morally loaded) framing and shaming in a 2×3 factorial design. Our experiment combines the disclosure of individual contributions plus visual identification if behavior fell short of what was described as socially warranted, with three different descriptions: first, in a neutral form as a voluntary contribution mechanism (VCM); second, as a morally loaded tax avoidance game where not paying taxes was framed as legally adequate but socially questionable behavior; and third, as a game presented in a tax (avoidance) context but without moral loading.

For framing and shaming as separate behavioral triggers, we confirm general findings from lab and field experiments on voluntary contributions and tax evasion (see the ‘Related literature’ section): shaming works best through blatant exposure and has a positive effect on prosocial behavior. Moral framing has a positive effect on tax compliance both in comparison with voluntary contributions and the simple tax frame, but fades quickly with no social sanctioning mechanism in place.

In our experiment, framing tax avoidance without moral loading led to equal compliance rates as in the baseline VCM, and in combination with disclosure, to significantly less contributions than in the morally loaded setting. Thus, we find evidence that tax avoidance itself has no implicit moral connotation; this had to be saliently introduced by describing it as morally ‘wrong’ behavior, exemplifying the importance of joint effects and interdependencies, which are scarcely studied and are the main novelty of our analysis. Adding to the above result, we show that with moral loading, the willingness to be revealed as a tax avoider is significantly smaller than in the case when the possibility of avoiding taxes is not combined with a moral appeal. Also, the effectiveness of framing and shaming was highly sensitive to good and bad example setters: without moral loading and/or disclosure, few participants were initially fully compliant, which led to a quick spreading of noncooperative behavior. Such social contagion also played a role in the disclosure treatments: in groups with few good example setters (high contributions in early rounds), compliance rapidly eroded. While this was a rare event with moral loading and disclosure, it happened frequently with the simple tax frame, where tax avoidance was not understood as a ‘wrong’ choice, or in other words, as failure of moral duty.

Related literature

Abandoning anonymity in economic games has increasingly attracted attention in recent years: Andreoni and Petrie (2004) used pictures to make donors and their contributions identifiable to fellow group members. Such visual identification led to significantly increased contributions to the public good. Interestingly, donations were the highest when disclosure was a deliberate choice. Samek and Sheremeta (2014) displayed pictures plus names of participants who contributed less than the maximum

possible amount to a public good. They observed significantly increased contributions resulting from this treatment. Rege and Telle (2004) conducted a one-shot public goods game where participants could constantly observe each other. This significantly increased cooperation, while a second treatment that framed cooperation as a social norm had no effect. Noussair and Tucker (2007) replicated Rege and Telle's experiment over 20 rounds and found that shaming depleted quickly. Bochet *et al.* (2006) dissolved anonymity by allowing face-to-face communication and chat boxes. Participants who met and talked before the actual experiment started made significantly larger contributions than the anonymous control group.

While our experiment is the first that studies disclosure in the context of legal tax deficiency, articles on illegal tax evasion point to a general usefulness of shaming: Blaufus *et al.* (2017) and Casal & Mittone (2016) find a positive effect of disclosure (via participants' pictures) on tax-declared income. However, detected tax evasion was monetarily sanctioned and compliance thus incentivized with a pecuniary instrument. Fortin *et al.* (2007) and Coricelli *et al.* (2010) find similar results in tax evasion games without the inclusion of a public good, and Coricelli *et al.* additionally show that cheating was accompanied by emotional arousal under the threat of shaming. Alm *et al.* (2017) distinguish between the decisions to participate in tax evasion and between the amount evaded. They find that shaming foremost deterred participation in tax evasion and that shaming worked regardless of differing social norms on compliance.

Even if not explicitly written in law or practically enforced, taxpaying implies a legal obligation, a necessary evil that one cannot (or should not) escape. In a lab experiment on taxpaying, Blaufus *et al.* (2016) show that an economically equivalent decision was treated differently when it was framed as (admissible) tax avoidance versus (illegal) tax evasion. In the evasion scenario, tax minimization was less pronounced, but the differences vanished once pecuniary consequences were introduced. Tyran and Feld (2006), in a lab experiment, observed that exogenously imposed penalties did not raise cooperation when they were nondeterrent.¹

Our disclosure treatments also connect to 'real-world' shaming practices: Dwenger and Treber (2018) analyze a recently implemented naming-and-shaming policy in Slovenia where tax delinquents (self-employed and corporations) were put on a publicly available list. The authors show that the threat of being placed on this list led to a reduction in tax debt by around 8.5%, whereas the effects of actually being shamed are comparably marginal. Bø *et al.* (2016) exploit a policy change in Norway in 2001, when information on tax returns became available on the Internet. Following this policy change, reported taxable income increased by 3%. Since the effect was the largest in densely populated areas, Bø *et al.* (2016) conclude that it was driven mainly by the wish to avoid media attention or public shaming. Perez-Truglia and Troiano (2015) find evidence for reduced tax evasion in the USA when tax delinquents were shamed by informing their neighbors.

The framing of a tax context, the second behavioral trigger in our experiment, has also found attention in economic games on taxation: Baldry (1986) framed gambling

¹For a local church tax in Bavaria with historically zero audits and fines, Dwenger *et al.* (2016) show that deterrence only modestly reduces tax evasion. Most tax payments seem to be driven by duty-to-comply preferences. However, endogenously imposed sanctions did increase prosocial behavior.

versus tax evasion and found significantly less gambling in the tax context. He interpreted this as the result of moral costs arising from the involvement in a cheating activity. Wartick *et al.* (1999) also found a tax reporting context to matter for decisions in a lab experiment. However, their result was driven by the finding that older participants reported twice as much income in the tax context than younger participants. Durham *et al.* (2014) found that compliance in their lab experiment did not depend on whether the available endowment was earned (and subsequently taxed) or not. Alm *et al.* (1992) found that cooperation in the lab was not affected by introducing a tax context. Instead of perceiving tax evasion as morally wrong, the authors see tax compliance as the consequence of overwhelming low (audit) probabilities.

The effectiveness of framing a 'moral duty' to comply with the tax code finds mixed support in the literature. In particular, field experiments on tax compliance reveal that simple normative appeals or stating a 'golden moral rule' are mostly ineffective and short-lived, compared with, for example, messages on penalty salience or others' compliance (Blumenthal *et al.*, 2001; Dal Bó & Dal Bó, 2014; Luttmer & Singhal, 2014; Hallsworth *et al.*, 2017; Meiselman, 2018). A smaller number of studies find positive effects of moral appeals, for example, in the context of compliance with a mandatory but poorly enforced public broadcasting fee in Austria (Fellner *et al.*, 2013) or letters with moral appeals sent to potential tax avoiders in Norway, which on average doubled self-reported foreign income (Bott *et al.*, 2017).

Combining (morally loaded) tax framing and disclosure of decisions, the focal point of interest in our experiment raises important questions about how contagious the observation and potential abundance of norm violations is, and whether the term tax itself might already entail a norm that participants followed. For example, Andreoni & Bernheim (2009) and Lazear *et al.* (2012) found that with the opportunity to conceal choices, generally, generous individuals behaved selfishly in dictator games. Kimbrough and Vostroknutov (2016) matched participants in a VCM according to their preferences for rule-following and found that cooperation swiftly decayed in low-preference groups, while it prevailed in those with a strong preference for rule adherence. In another dictator game, Dimant (2015) suggests that unethical behavior is more contagious than observing good example setters. Gino *et al.* (2009) analyzed the reaction to observing unethical behavior with a psychological experiment. They find that copying a norm-violation (cheating on a test) was crowded-in when the culprit was an in-group member, and crowded-out otherwise.

The experiment

The 2×3 factorial design of our public goods game is sketched in Table 1; it allows not only to compare each state of framing (*Baseline*, *Tax frame*, and *Moral frame*) with the effect of introducing disclosure, but also to analyze interactions within and across framing and disclosure scenarios, for example, between *Tax shaming* and *Moral shaming* or between *Moral frame* and *Tax shaming*.

Each of the six variants drew on an identical contribution mechanism and payoff function: participants were provided with an endowment of $E = 100$ Experimental Currency Units (ECU). Of these 100 ECU, between 0 and 40 ECU could be invested

Table 1. The experimental design.

	Disclosure	
	No	Yes
Framing		
No	Baseline	Shaming
Tax	Tax frame	Tax shaming
Loaded	Moral frame	Moral shaming

into a productive public venture.² The remaining ECU went to a private account with a return rate of 1. The experimental payoffs for player i were $\pi_i = E - x_i + g_i$, where $x_i \in [0, 40]$ denotes the contribution to the public venture by player i and g_i is i 's payoff from the public account. This return is given by $g_i = (\gamma/N) (x_i + X_{-i})$, where X_{-i} is the sum of investments to the public good by the (four) other group members except player i , and γ is an efficiency factor (rate of return) of the public good; we chose $\gamma = 1.5$ in all treatments. The parameters described here resulted in the payoff matrix depicted in the instructions (can be found in the Supplementary Material).

Groups consisted of $N = 5$ members and were fixed in composition over the 10 rounds of the experiment. We used fixed composition to introduce familiarity (of members and their behavior) and a sense of publicity within groups, which would not have been possible with random matching. With fixed group composition, members learn about their fellow participants' behavior, and in the disclosure treatments, they can even identify individual contributions. Such information will likely result in adaptation, which could lead either to a reinforcement of an in-group (tax) compliance norm or to the opposite if widespread free-riding is observed: when participants observe free-riding in their group, they might also not invest in the profitable public good. This social contagion should be strongest in groups where noncompliance violates the moral code we framed in the experiment. On the other hand, observing widespread compliance could also result in a prosocial outcome, which should also be facilitated by moral framing and disclosure of noncompliers as a social control.

Within standard-economic logic, a VCM (like in our Baseline) should result, in a finite game, in a Nash equilibrium of zero contributions to the public good by each group member in every round. Since it is well-established that players in a VCM behave at least partially cooperative, we expect the same in our experiment. The treatments study (and combine) different nonpecuniary incentives to behave more cooperatively, compared with the simple VCM. These incentives are not payoff-relevant, and, thus, exclusively behavioral triggers with a foundation in the behavioral economic literature (see the above section). We expect the treatments to work through different channels and to be of different effectiveness, as laid out in more detail below.

²The choice was discrete and made in steps of 5 ECU, reflecting different tax rates. We use the 60:40 split to simulate a (maximum) tax rate of 40% in our framed experiments.

Tax framing

For *Tax frame* and *Tax shaming*, we altered the wording of the public goods game in the instructions. We replaced ‘endowment’ with ‘taxable income’ and ‘contribution’ with ‘tax payments’. The neutral option to invest in a profitable public project was changed into a linear income tax that would finance a public project. Individuals had the option to reduce their personal tax burden by choosing any lower tax rate than stipulated by the law. We explicitly mentioned that this possibility was not accompanied by any monetary punishments.

An interesting question we can address with these treatments is whether a specific morality is *per se* connected to tax avoidance or whether it has to be called into attention. Thus, we are able to disentangle the simple tax frame from moral loading.

Moral loading

To saliently introduce that taxes are actually meant to be paid and to make participants consider the moral dimension of tax avoidance, we presented a short text on-screen, laid out as a newspaper commentary in the treatments *Moral frame* and *Moral shaming*.³ Under the headline ‘Tax avoidance is legal, but can it ever be legitimate?’, the text briefly defined tax avoidance as a legal way to reduce one’s personal tax burden that might still not have been intended by the government. Thus, we insinuated that tax avoiders would not act in the spirit of the law, even if they technically did not violate it. Toward the end, we reminded readers that tax avoidance is a personal decision; whether it was seen as socially justifiable was deliberately left to participants’ own judgment.⁴

We expect that, related to field experiments by Bott *et al.* (2017) and Hallsworth *et al.* (2017), framing a specific morality and a duty of paying taxes will have a positive effect on cooperation over and above the Baseline and the simple tax frame, even if compliance can be enforced neither by pecuniary consequences nor by reputational consequences. However, with no sanctions to prevent self-interest in place, we figure that the positive effect on contributions will not last long, mainly due to observing (via the size of the group account) and copying others’ noncompliance.

Disclosure

In the experiment, information on the own investment, the sum of the investments in the group, and the individual payoff in the respective round were presented to participants in all six variants of the game. In the disclosure treatments, this information was followed by the photographs and the actual investment of those group members who contributed less than the maximum possible share of their endowment. Full contributions meant complete anonymity. That is, group members who did not once free-ride were never revealed to their fellow group members.⁵ If all group members

³The text was written by us, presenting a representative view of commentaries in German media following the Panama Papers leak. Showing an actual commentary would not have satisfied these requirements, and could have even biased behavior by a specific political view of the media outlet.

⁴Supplementary Figure B shows the newspaper commentary in German as well as its English translation.

⁵Supplementary Figure C shows an example of the disclosure screen.

chose full compliance, a note reporting this outcome was displayed on-screen. Since with disclosure social punishment occurs even after the slightest violation of behaving socially optimal, we predict that many participants do not want to be revealed as rule breakers, and that this will be the case regardless of the context in which the disclosure of photographs and individual decisions occurs.

Besides the general effectiveness of shaming practices (see the ‘Related literature’ section), with our experimental design, we expect to find qualitative differences in shaming, which depend on moral loading. If certain behaviors and actions are perceived as ‘normal’ or ‘ethical’, or at least when they are not explicitly described as socially harmful, the experience of shaming might not go beyond a general aversion to be drawn into the center of attention and to leave the comfort of anonymity. If, on the other hand, engagement in a certain activity (here, tax avoidance) is explicitly described as harmful, this moral loading might induce an additional disincentive to be revealed as an opportunistic cheap-profitteer.

Experimental protocol and summary statistics

The experiment was conducted at the computerized laboratory (LLEW) at the Leibniz University Hannover in August and September 2017 and for additional data collection (conditions *Tax frame* and *Tax shaming*) in February 2019.⁶ Participants were recruited from the general student population with the software hroot (Bock *et al.*, 2014). A total of 300 subjects (154 males, 145 females, and 1 subject made no statement) participated in the experiment. Earnings averaged around 11 Euro in approximately 1 h. Additionally, participants received a lump-sum showup fee of 4 Euro. Of the 300 participants, 55 played in each of the *Shaming*, *Moral frame*, and *Moral shaming* treatments (4 sessions and 11 groups per treatment). Fifty participated in the Baseline (4 sessions and 10 groups), 45 in the *Tax shaming* condition, and 40 in the simple *Tax frame* treatment. The experiment was programmed with z-Tree (Fischbacher, 2007).

Before the start of a session, one treatment was randomly selected; subjects were randomly seated and then matched (according to their seat number) to groups of five by the experimental software. The participants’ photos were taken right before the instructions (see Supplementary Material) were handed out. To avoid differential expectational effects, we also took photos in the treatments where decisions would not be disclosed. Participants gave written consent to shortly saving and potentially using their photo in the experiment (see Supplementary Material). If a subject would not sign the consent form, he or she was not permitted to participate but still received the showup fee. This happened only in one case. Before playing the experiment, participants had to answer a short computer-based comprehension test (see Supplementary Material). In the morally loaded treatments, the newspaper comment was presented to the subjects on-screen for two and a half minutes before the first decision was made. After the last decision had been made in round 10, the payoff of one round was randomly selected and paid in cash to the participants.

⁶Participants in 2017 and early 2019 were recruited from the same pool of students and the recruitment followed exactly the same procedure. We do not find systematic differences between the data collected in 2017 and 2019 (see Table A1). In the multivariate analysis, we additionally control for potential influences of socioeconomic characteristics.

Table 2. Summary statistics of socioeconomic characteristics.

Variable	Mean	Median	Standard deviation
Female	48.5%		
Economics Major	20.0%		
Bachelor degree	38.3%		
Employed	34.7%		
Tax declaration	61.7%		
Age	24.21	24.00	4.23
Income	347.46	300.00	254.28
Semester	7.18	7.00	4.02

Notes: The total number of subjects is 300. 'Economics Major' indicates whether a subject studies economics or management. 'Bachelor degree' takes the value of 1 if it is the subject's highest educational degree. 'Employed' indicates whether a participant holds a job besides studying. 'Tax declaration' takes the value of 1 if the subject has at least once in life filed a tax declaration. 'Income' is monthly disposable income after deducting all fixed expenses. See Table A1 for summary statistics separated by treatment.

Before payouts, we asked participants to answer a short socioeconomic questionnaire. The subjects' socio-demographics are summarized in Table 2: Our sample is quite balanced between female and male participants. A relatively low share of 20% of participants was enrolled in an economics major program. Another 20% of the sample studied at the department of philosophy, 22% in an engineering-related major, and 12% were enrolled in a natural science program. With these numbers, our sample represents quite a good cross section of the student population in Hannover.

Table A1 additionally shows socio-demographics for each of the six experimental conditions. For some treatments, there are slight differences, for example with respect to Bachelor degrees or taxpaying experience; however, they had no impact on the investment decisions in the experiment (see the 'Multivariate regression analysis' section).

Results

Descriptive statistics and nonparametric sample tests

Panel A of Table 3 reports the average contributions to the public good (in ECU) over all 10 rounds of the experiment. Regardless of the framing context, disclosure has a strong impact on cooperative behavior. Surprisingly, disclosure in the neutral and morally loaded tax avoidance treatments results in equivalent contributions, while *Tax shaming* yields compliance one third below *Moral shaming*. Differences between *Tax frame* and the Baseline VCM are negligible, but introducing the moralizing newspaper comment results in tax payments 5 ECU above the simple tax frame. Across the framing stages, introducing disclosure has the largest impact in the neutral Baseline, where cooperative behavior is entirely voluntary.

Panel B of Table 3 reports nonparametric sample tests to analyze the statistical significance of this first impression. Not all possible combinations of treatments are

Table 3. Average contributions to the public good and nonparametric sample tests.

		Disclosure	
		No	Yes
Panel A: Average contributions to the public good/tax payments in ECU			
Framing	No	13.8 (14.09)	31.32 (14.34)
	Tax	14.2 (15.19)	22.23 (17.37)
	Loaded	19.25 (16.32)	31.11 (15.09)
Panel B: Nonparametric Mann-Whitney U-tests			
Rounds	First	2–10	Last 3
Level	Individual	Group	Group
Baseline	19.10 (15.11)	13.21 (9.38)	9.40 (8.76)
Tax frame	21.25 (16.20)	13.42 (8.34)	9.33 (7.45)
Moral frame	27.64 (15.33)***	18.32 (10.77)***	12.00 (7.75)***
Shaming	35.27 (10.69)***	30.88 (10.63)***	24.48 (12.45)***
Tax frame	21.25 (16.20)	13.21 (9.38)	9.33 (7.45)
Moral frame	27.64 (15.33)*	18.32 (10.77)***	12.00 (7.75)***
Shaming	35.27 (10.69)***	30.88 (10.63)***	24.48 (12.45)***
Tax shaming	32.56 (11.90)***	21.09 (12.90)***	16.48 (13.51)***
Moral frame	27.64 (15.33)	18.32 (10.77)	12.00 (7.75)
Shaming	35.27 (10.69)***	30.88 (10.63)***	24.48 (12.45)***
Tax shaming	32.56 (11.90)*	21.09 (12.90)***	16.48 (13.51)**
Moral shaming	34.27 (12.00)***	30.76 (10.79)***	25.76 (12.42)***
Tax shaming	32.56 (11.90)	21.09 (12.90)	16.48 (13.51)
Moral shaming	34.27 (12.00)	30.76 (10.79)***	25.76 (12.42)***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Notes: Panel B shows average contributions to the public good (in ECU) on the individual level (first column) and the group level (second and third column). Standard deviations are in parentheses. Statistical significance of the differences in contributions, relative to a reference treatment (bold numbers), is tested with a two-sided Mann-Whitney U-test.

tested, because, for example, Baseline and *Moral shaming* differ in framing, moral loading, and disclosure, which cannot be disentangled at this point of the analysis. Due to the fixed group-matching, participants' decisions in the experiment are independent of others' behavior only in the first round. To account for this, Panel B of Table 3 shows individual contributions only for the initial decision. For the subsequent rounds (2–10), the analysis is on the group level, that is, the size of the group account, divided by five.

In the experiment's first round, participants in the Baseline and the *Tax frame* treatment contributed roughly half of what they could to the public account. Compliance in the disclosure treatments was very high with a rate of 82–88%. In the first round, *Tax shaming* and *Moral shaming* were statistically not different.

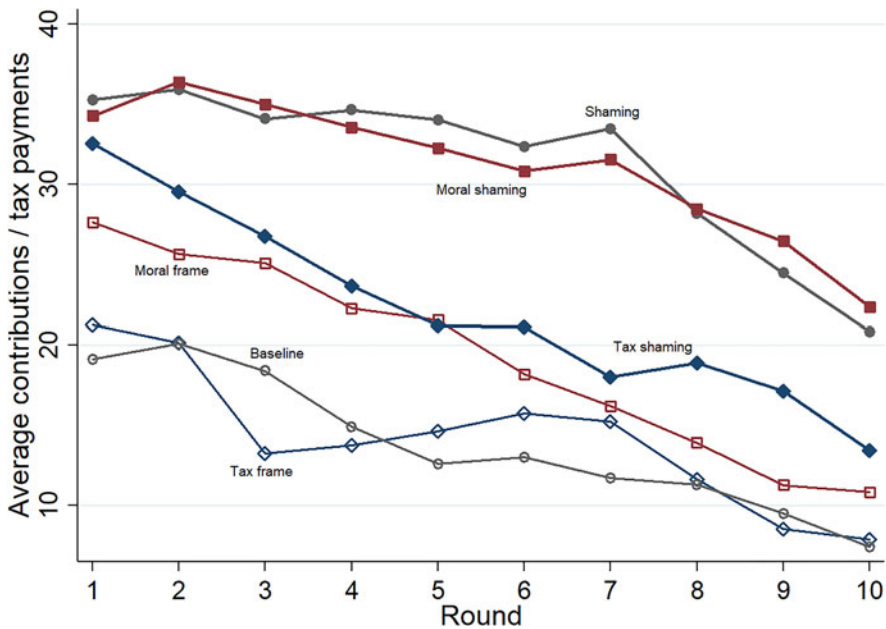


Figure 1. Contributions (tax payments) over time, in ECU.

This changed (on the group level) in subsequent rounds, where compliance with *Tax shaming* decreased much more rapidly than in the morally loaded disclosure treatment. With morally loaded framing, participants contributed about 70% of what they could in the first round. The 27.6 ECU were still significantly below the disclosure treatments, but significantly above the tax avoidance setting without moral loading. For rounds 2–10, the average contributions on the group level closely resemble Panel A of Table 3. Although cooperative behavior decreased over time, statistical significance across treatments remains robust throughout all three levels of analysis (thus, also when only the last three rounds are considered). Moral framing always remains significantly above tax framing and contributions in the *Shaming* and *Moral shaming* treatments are still around 75% in the last late stage, when last round effects might kick in.

Figure 1 visually reinforces the results of Table 3: contributions with *Tax Shaming* start to decline as early as in the second round, while they remain high and very steady until round eight in the other two disclosure treatments. Interestingly, the *Moral frame* treatment has a very similar dynamic as *Tax shaming*, only on a different starting level. In the last round, contributions in the latter two treatments fall close to 10 ECU, only slightly above the Baseline and the *Tax frame* treatment.

The treatments *Tax shaming* and *Moral frame* also share the characteristic of few participants contributing 40 ECU to the public good in most (if not all) of the experiment’s 10 rounds and a majority of subjects who very infrequently behaved fully compliant – mostly in the first couple of rounds. Table 4 shows that this is in stark contrast to *Shaming* and *Moral shaming*, where 31–38% invested 40 ECU in all 10

Table 4. Frequency of full compliance in different round intervals.

Rounds with full compliance (40 ECU)	Baseline	Tax frame	Moral frame	Shaming	Tax shaming	Moral shaming
10 rounds	6%	0	1.82%	30.91%	15.56%	38.18%
7–9 rounds	2%	7.5%	5.45%	27.27%	20%	20%
3–6 rounds	6%	20%	36.36%	25.45%	17.78%	27.27%
0–2 rounds	86%	72.5%	56.36%	16.36%	46.67%	14.55%
<i>Total observations</i>	<i>N = 500</i>	<i>N = 400</i>	<i>N = 550</i>	<i>N = 550</i>	<i>N = 450</i>	<i>N = 550</i>

Notes: The numbers represent full-compliance decisions (investing 40 ECU into the public good) as a percentage of total observations per treatment for four different round intervals, ranging from (almost) no compliance (0–2 rounds) to always full compliance (10 rounds). Nonparametric sample tests for statistical differences across treatments can be found in Table A2.

Table 5. Good and bad example setters.

	Baseline	Tax Frame	Moral frame	Shaming	Tax shaming	Moral shaming
Panel A: Distribution of good and bad example setters (% of subjects)						
Good example	10.00	7.50	25.45	61.82	37.78	63.64
Bad example	44.00	50.00	18.18	5.45	11.11	5.45
Panel B: Share of groups with a majority of good or bad example setters						
Good example	10	12.5	54.55	90.91	55.56	90.91
Bad example	80	87.5	36.36	9.09	11.11	0
Panel C: Frequency of full compliance by followers (in rounds 4–10)						
Good example	4.76	42.86	20	45.92	38.57	52.04
Bad example	1.68	14.29	3.30	14.29	0	0
<i>p</i> -value	0.546	0.009	0.001	0.003	<0.001	<0.001

Notes: Good example setter: contributions of 40 ECU in each of the first three rounds of the experiment. Bad example setter: a total of 45 ECU invested into the public good over the first three rounds. Followers are subjects who were neither identified as good nor as bad example setters. If the numbers in Panel B do not add up to 100%, good and bad example setters were equally distributed. Significance across classifications in Panel C is tested with the Kruskal–Wallis test.

rounds (and, thus, never revealed their pictures to fellow group members), while almost 60% of participants fully contributed in at least seven rounds.

This is significantly above (Mann–Whitney *U*-test, see Table A2)⁷ the 15.6% (all 10 rounds) respectively 20% (7–9 rounds) with *Tax shaming*; almost half of the participants in this treatment contributed 40 ECU in no more than two rounds.

⁷Table A2 provides nonparametric sample tests for the different full-compliance categories across treatments in the style of Table 3, Panel B. Almost all percentage-differences are statistically highly significant (Mann–Whitney *U*-test).

Permanent or frequent full compliance was almost nonexistent in the Baseline and with the simple *Tax frame*. In particular with morally loaded tax avoidance, infrequent compliance (3–6 rounds) was common, and 56% of subjects behaved socially optimal only in at best (the first) two rounds. Whenever disclosure was involved, frequent or permanent compliance significantly increased, pointing to a positive effect of this social sanction on average contributions (Table 3) and also on the probability of foregoing personal gains in favor of anonymity (Table 4).

Good and bad example setters

A solid and growing body of literature argues that the copying and adopting of (perceived) social norms and examples set by others can explain behavioral patterns like those we observed in our experiment (thus, different responses to disclosure, dependent on the respective framing). Indeed, as Panel A of Table 5 shows, treatments with high average contributions and a high number of (at least) frequent compliers (see Table 4) have a large share of *good example setters*, which we define here as a subject fully contributing (40 ECU) to the public good in each of the experiment's first three rounds (these early rounds arguably act as a stage of strategy formation, for example by observing others' decisions). The exact opposite is true for the nondisclosure scenarios Baseline and *Tax frame*, where the majority of subjects set a 'bad example' by contributing less than 45 ECU over the first three rounds of the experiment.⁸ The distribution of good and bad example setters under morally loaded framing and the observation that more than half of participants had no clear strategy in the first three rounds in this treatment corroborate previous arguments: with the *Moral frame*, participants were initially intrigued by our comments on tax avoidance, but likely tempted to violate this call due to the lack of (social) consequences.

The result of this abeyance was that, as depicted in Panel B of Table 5, 55% of groups in the *Moral frame* setting were led by a majority of good example setters, while 36% were outranked by low- or never compliers. The picture in the other treatments is clear cut: with *Shaming* and *Moral shaming*, 91% of groups had a majority of good example setters; the opposite was the case for the Baseline and the simple tax frame. *Tax shaming* closely resembles *Moral frame* in terms of good example setters, but the other disclosure treatments with respect to the share of bad example setters.

In Panel C of Table 5, we label participants who were neither good nor bad example setters as 'Followers' and check whether their behavior (investing 40 ECU into the public good or less) was affected by the distribution of positive or negative role models, adding information to the dynamics in Figure 1.

Indeed, we find statistically significant differences in Followers' compliance: confronted with a majority of good example setters, 40–50% of decisions (in rounds 4–10) in the disclosure treatments were to contribute the maximum possible 40 ECU. In groups with a majority of bad example setters, the frequency dropped to

⁸The cutoff value for bad example setters is oriented at average contributions (in the first three rounds) in the treatments without any moral or disclosure-related behavioral incentives. Nevertheless, the cutoff is to some extent arbitrary and can separate comparable groups into different categories (more good vs. a majority of bad example setters).

14% (*Shaming*) and even zero with *Moral shaming* and *Tax shaming*. The simple *Tax frame* treatment is an interesting and special case: in the (very few) groups with more good than bad example setters, 43% of all investment decisions were full compliance, more than doubling the *Moral frame* outcome. Even with a majority of bad example setters, full compliance was markedly more common than in the Baseline and with moral framing.

Multivariate analysis

In this section, we report regression analyses for our dependent variable *Contribution*, analyzing treatment effects with linear random effects and mixed-effects regressions with the subject ID as the cross-sectional variable and the round number as the time variable. The predictors of interest are the treatment dummy variables (see Table 6), taking a value of 1 if a participant was assigned to a respective treatment and 0 else. Effects are measured as differences in ECU against the reference category Baseline. Socioeconomic characteristics (see Table 2) and a continuous round measure are included as control variables.

With the regression analysis, we specifically address intragroup dependencies pointed out in the preceding paragraphs: first, with the mixed-effects models, we include the independent groups as random intercepts. Second, standard errors are clustered on the group level in all specifications in Table 6. Third, we introduce a dummy variable in Table 6 that takes (and holds) a value of 1 after the first time a participant contributed less than 40 ECU to the public good. Thus, for a group member who fully complied in the first round and contributed, for example, 30 ECU in the second round, the dummy 'Revealed' takes a value of 1 in the following seven rounds of the experiment. With this dummy we capture whether the experience of having been disclosed and 'shamed' has an effect on participants' subsequent behavior, thus, if the aversion of being revealed is smaller once a participant has been confronted with disclosure, resulting in lower contributions. Lastly, the treatment variables are interacted with the continuous round measure to check whether (negative) effects of progressing time on contributions are treatment specific.

Column (1) of Table 6 reports the results of the linear random effects regression over all 10 rounds and generally confirms the results from the nonparametric tests after controlling for socioeconomic characteristics and the time passed in the experiment. The simple tax frame leads to contributions similar to the Baseline. The other treatment effects are significant and the coefficient sizes are comparable to the differences reported in Panel B of Table 3. From Table 6, we cannot directly deduce the differential treatment effects. However, this is easily done by subtracting, for example, *Moral frame* from *Moral shaming*, thus $16.4 \text{ ECU} - 5.5 \text{ ECU} = 10.9 \text{ ECU}$. A postestimation Wald test shows that the difference from adding disclosure to moral framing is statistically highly significant ($p = 0.002$, not reported in Table 6) and tops the additional impact of shaming on the simple tax frame by around 2.5 ECU; this difference is not very large, but the treatments *Moral shaming* and *Tax frame* differ significantly on the 10% level (Wald test, Table 6); interestingly, the 3 ECU difference between the *Moral frame* and *Tax shaming* condition is not significant in the multivariate context ($p = 0.463$, Wald test).

Table 6. Linear random and mixed-effects regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RE	ME	RE	ME	RE	ME	RE	ME
Rounds	All 10	All 10	All 10	All 10	All 10	All 10	Last 3	Last 3
<i>Interaction</i>			Revealed	Revealed	Round	Round		
Tax frame	0.014 (2.811)	0.168 (2.940)	-1.439 (3.405)	-2.603 (3.854)	-0.678 (3.135)	-0.524 (3.235)	-0.591 (3.305)	-0.303 (3.430)
Moral frame	5.457* (3.226)	5.198 (3.355)	7.299** (3.485)	4.295 (3.780)	9.212*** (3.542)	8.953** (3.662)	2.250 (3.123)	2.287 (3.268)
Shaming	17.10*** (3.153)	16.53*** (3.403)	17.06*** (3.079)	12.98*** (3.441)	17.96*** (2.401)	17.38*** (2.623)	14.44*** (4.127)	13.81*** (4.358)
Tax shaming	8.435** (3.918)	7.876** (3.997)	11.04*** (3.588)	8.023** (3.523)	10.74*** (2.751)	10.18*** (2.821)	7.116 (4.809)	6.592 (4.912)
Moral shaming	16.35*** (3.309)	16.91*** (3.438)	16.25*** (3.058)	13.03*** (3.409)	16.40*** (2.539)	16.95*** (2.652)	14.90*** (4.321)	15.66*** (4.382)
Revealed			-4.105** (1.624)	-8.887*** (1.841)				
Round	-1.563*** (0.138)	-1.563*** (0.138)	-1.122*** (0.138)	-0.988*** (0.138)	-1.364*** (0.204)	-1.364*** (0.204)	-2.509*** (0.458)	-2.509*** (0.454)
<i>Interactions</i>								
Tax frame			1.819 (2.228)	3.522 (2.442)	0.126 (0.417)	0.126 (0.415)		
Moral frame			-3.001 (2.188)	0.343 (2.291)	-0.683** (0.291)	-0.683** (0.290)		
Shaming			-4.161 (3.049)	0.201 (2.973)	-0.156 (0.364)	-0.156 (0.362)		
Tax shaming			-5.473 (3.473)	-2.704 (3.156)	-0.418 (0.464)	-0.418 (0.462)		
Moral shaming			-3.464 (2.970)	0.873 (2.492)	-0.008 (0.471)	-0.008 (0.469)		
Age	0.358* (0.188)	0.0733 (0.143)	0.313* (0.167)	0.0614 (0.126)	0.358* (0.188)	0.0732 (0.143)	0.381 (0.323)	0.0733 (0.287)
Female	-1.433 (1.239)	0.449 (0.916)	-0.892 (1.072)	0.616 (0.752)	-1.433 (1.240)	0.450 (0.916)	-0.253 (1.631)	1.732 (1.254)
Semester	-0.0328 (0.189)	-0.210 (0.164)	-0.0880 (0.168)	-0.249* (0.133)	-0.0328 (0.189)	-0.210 (0.164)	0.0805 (0.236)	-0.181 (0.211)
Econ. Major	-4.485** (1.774)	-4.233*** (1.441)	-4.356*** (1.511)	-4.044*** (1.257)	-4.485** (1.775)	-4.233*** (1.441)	-4.749** (2.281)	-4.930*** (1.805)
Bachelor	0.690 (1.288)	0.959 (1.276)	0.850 (1.087)	1.315 (1.025)	0.690 (1.289)	0.959 (1.276)	-0.212 (1.611)	0.763 (1.549)

(Continued)

Table 6. (Continued.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RE	ME	RE	ME	RE	ME	RE	ME
Rounds	All 10	All 10	All 10	All 10	All 10	All 10	Last 3	Last 3
Interaction			Revealed	Revealed	Round	Round		
Income	-0.007*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.005*** (0.002)	-0.007*** (0.002)	-0.006*** (0.002)	-0.007** (0.003)	-0.007*** (0.002)
Employment	-2.924** (1.244)	-1.833* (1.088)	-2.328** (1.060)	-1.198 (0.983)	-2.924** (1.245)	-1.832* (1.088)	-2.431 (1.630)	-1.406 (1.392)
Tax experience	-0.929 (1.128)	-1.667* (0.955)	-0.759 (0.975)	-1.218 (0.743)	-0.929 (1.129)	-1.667* (0.955)	-2.069 (1.499)	-3.031** (1.344)
Observations	2810	2810	2810	2810	2810	2810	843	843
Number of subjects	281		281		281		281	
R ² -within	0.159		0.161		0.164		0.056	
R ² -between	0.357		0.515		0.357		0.239	
R ² -overall	0.269		0.355		0.272		0.193	
Wald tests (p=)								
Tax frame = Moral frame	0.074	0.112	0.003	0.039	0.009	0.014	0.38	0.442
Tax shaming = Moral shaming	0.051	0.03	0.048	0.028	0.012	0.003	0.159	0.108
Moral frame = Tax shaming	0.463	0.518	0.227	0.189	0.653	0.722	0.306	0.378
Tax frame = Tax shaming	0.023	0.044	<0.001	<0.001	<0.001	<0.001	0.108	0.166

*p < 0.1; **p < 0.05; ***p < 0.01.

RE = Random effects; ME = Mixed effects.

Notes: The dependent variable in all specifications is the contribution to the public good in ECU. The reference category for the treatment effects is the Baseline. See Table 2 for a description of the socioeconomic variables used in the regression analysis. Standard errors, clustered on the group level, are in parentheses.

Column (2) specifies a mixed-effects model of column (1). The coefficients are slightly corrected in their magnitude, but except for *Moral frame*, the statistical significance of the treatment variables is unaffected; this holds when random- and mixed-effects models are estimated for the experiment's last three rounds (columns (7) and (8)). There, only the effects of *Shaming* and *Moral shaming* remain significantly above Baseline.

Columns (3) and (4) introduce the 'Revealed'-indicator. Overall, contributions respectively tax payments after the first decision to invest less than 40 ECU are significantly lower than in previous rounds. This negative effect is much more pronounced in the mixed-effects model, fortifying the importance of the in-group dynamics presented in Table 5. However, the interaction coefficients are not significant for the single experimental conditions, although the disclosure treatments only just miss statistical significance at the 10% level. Including the Revealed-treatment interactions does not affect the impact of *Shaming* and *Moral shaming* (over Baseline) in the RE model but reduces their impact in column (4). *Tax shaming* and *Moral frame* in column (4) are quite similar to the ME model in column (2).

Table 7 can give some additional reasoning behind these findings: first, average contributions in the *Shaming* and *Moral shaming* treatments at the first noncompliance fell close to the Baseline level, but rebounded to above 50% in the rounds after having been revealed for the first time. This is not at all the case with *Tax frame* and *Moral frame*, where cooperation remained equivalent to the contribution at the first noncompliance. In the third disclosure treatment, *Tax shaming*, average contributions after the first noncompliance decrease by around 4 ECU, compared with the contributions at the first noncompliance. Relating to Table 5, these findings could be driven by the different number of good and bad example setters, who act as role models to Followers: with *Shaming* and *Moral shaming*, 34% respectively 39% of all decisions (after the first noncompliance) were to invest 40 ECU, which is much higher than the 21% with *Tax shaming* and vastly above our *Tax frame* and *Moral frame* treatments. Table 7 also shows that participants in the *Shaming* and *Moral shaming* treatments played on average 4.4 respectively 3.4 rounds before they were first revealed to their fellow group members, about one round more than with *Tax shaming* and *Moral framing*.

Returning to Table 6, columns (5) and (6) report the results of our round-treatment interactions. Playing an additional round had a statistically significant negative effect on contributions in all specifications, and especially in the experiment's last three rounds (columns (7) and (8), statistical significance and effect magnitudes are close to column (3) in Panel B of Table 3). However, only the *Moral frame*-round interaction is significant, and controlling for such interdependencies increased the 'pure' impact of morally loaded tax avoidance on cooperation, relative to the Baseline.

Throughout all models in Table 6, only a few covariates have a significant impact on cooperation in our experiment. Students enrolled in an *Economics Major* were less cooperative, and the same generally held for participants who were employed (besides their studies). We found that *Income* had almost no impact on contributions, whereas older participants were slightly more compliant.

The results of our experiment have shown that, besides the amount contributed to the public good, disclosure is also connected to the likelihood of being fully compliant or not

Table 7. Behavior of shamed participants.

	Average contribution at first noncompliance (ECU)	Average contributions after first noncompliance (ECU)	Average round of first noncompliance	Frequency of full compliance after first noncompliance (%)
Baseline	14.36	10.77	1.34	3.93
Tax frame	12.00	12.29	1.70	10.84
Moral frame	14.91	15.29	2.46	11.79
Shaming	17.37	21.62	4.39	34.27
Tax shaming	18.55	14.27	2.66	20.79
Moral shaming	14.26	22.23	3.35	38.94

Note: Non-compliance is defined as a contribution of less than 40 ECU.

(see Table 4). This is understandable, since full compliance is tantamount to keeping an untarnished reputation, and the decision on how much tax to pay could be predated by a binary decision whether to avoid or not (see Alm *et al.*, 2017 for a related approach). In the Supplementary Material, we report results from nonlinear probit regressions on the likelihood of being fully compliant. There, we find that prosocial behavior on the intensive margin and the extensive margin are quite similar in our experiment. Almost 70% of decisions in the *Shaming* and *Moral shaming* treatments resulted in full compliance, compared with 42% with Tax shaming and only 25% with moral framing.

Discussion

Tax avoidance is, in many cases, sophisticated and legal tax planning, but overwhelmingly described as illegitimate, antisocial, and morally reprehensible by the media, politicians, and international organizations. In particular, prominent tax avoiders were named and publicly shamed for their actions amid the Panama and Paradise Papers revelations. With our experiment, we aimed to analyze causal links of disclosing individuals' participation in a legal yet (framed as a) morally questionable action – that is, tax avoidance.

One important takeaway is that tax avoidance in the experiment was not perceived as *per se* unethical by participants. When we saliently introduced moral loading to tax avoidance with our mock newspaper commentary, contributions to the public good (on average, over all 10 rounds) increased significantly above *Tax frame* and Baseline. To mirror extensive 'real-life' media reports on major tax leaks in recent years, we opted for a framing intended to trigger deliberation about the personal judgment and feelings toward tax avoidance, providing a transparent definition and a pro/contra argumentation. Thus, we decided against a simple behavioral rule like 'paying all of your taxes is the right thing to do', whose effectiveness finds only mixed support in the literature and what we thought to be likely too crude to convincingly convey the social consequences of tax avoidance – the lack of an effect with the simple *Tax frame* does us justice in this respect.

Shaming as a very general but well-suited definition is the action of publicly disclosing that a person (or firm, institution, etc.) behaved in a 'bad' way. In our experiment, this took place by disclosing pictures and individual tax payments, while 'bad' behavior was socially inappropriate tax avoidance. Under *Moral shaming*, where all these elements play together, we find a strong and temporally relatively stable positive effect, compared with the *Tax frame* and the *Moral frame* treatments.

The positive effect of *Tax shaming* on prosocial behavior is significantly less pronounced. With respect to the finding that tax avoidance was not *per se* perceived as illegitimate, this result appears to be reasonable: when a disclosed action is not (widely) perceived as 'bad' (because enough people engage in it), it is less shame-worthy than it was with *Moral shaming*. Then, higher contributions and especially a greater share of fully compliant participants might be rather the expression of general discomfort to being revealed to others and to become identifiable as a person that puts self-interest over society.

Interestingly and unexpectedly, shaming in the neutral context seems to be at play and as strong as in the morally loaded context. A number of studies, we referred to for example Andreoni and Petrie (2004), Samek and Sheremeta (2014), and Bochet *et al.*

(2006) in the literature overview, have demonstrated that disclosure can be an effective incentive for prosocial behavior in nonframed games or in a charitable giving scenario. Perhaps, in our experiment, participants in the *Shaming* treatment were very much aware of the social damage inherent to free-riding and cooperation connected to a specific social pressure to contribute, while failing to do so was perceived as something worthy to be ashamed of, and more compelling than dodging a tax that a few others were apparently paying. When no disclosure and shaming loomed (Baseline), however, there was no moral incentive to comply as a (weak) substitute.

In the experiment's first round, when participants had no information on others' behavior and attitudes toward free-riding, *Tax shaming* was statistically not different from *Moral shaming* and *Shaming*, and also tax compliance in the *Moral frame* treatment was considerable. It is, thus, important to understand what led to the swift decrease in cooperativeness in the two latter treatments, and the literature suggests contagion by copying others' behavior (Andreoni & Petrie, 2004; Gino *et al.*, 2009, Dimant, 2015) as an important contributor. We argue that the breakdown (or the preservation) of cooperation was indeed driven by these factors, and, in particular, by the number of good or bad example setters within groups (see Table 5): with *Shaming* and *Moral shaming*, the number of full- or frequent compliers overall and especially in the first three rounds was very high, resulting in the following suit of many low-compliance and undecided participants. This was much less the case with *Tax shaming*, where group members were often torn between complying and defying. With a large enough number of bad example setters, tax avoidance was not widely regarded as something 'wrong' or 'bad', that is, something shame-worthy. With moral framing, bad example setters gave the impression that there was not a shared understanding toward the illegitimacy of tax avoidance, and even for highly motivated group members, morals likely depleted.⁹

On a closing note, budget and organizational constraints limited the sample size particularly for the additional data collection in early 2019. Nevertheless, sound differences in contributions (especially when comparing nondisclosure with disclosure settings) are accompanied by high statistical power. This is especially the case in the experiment's first round with independent individual decisions, but to a slightly lesser extent also in rounds 2–10, when only the groups of five are truly independent observations (see Table 3). On the group level, comparing for example *Tax frame* versus *Moral frame* or *Tax shaming* versus *Moral shaming* has less statistical power. Yet, as we argue in the 'Discussion' and throughout the results section, diminishing differences between certain treatments and high volatility in cooperativeness are a result of observing others' behavior, contagion, and adaptation (see, e.g., Figure 1 and Table 5) and, thus, do not undermine our findings and their interpretation.

Concluding remarks

Our experiment interacts with two mechanisms that, in principle, help promote prosocial behavior when pecuniary incentives (rewards or fines) or legal enforcement are not available: moral framing and shaming. We, first, show that tax avoidance has *per*

⁹Wang *et al.* (2017) report interesting results of three studies on (priming) moral motivation and connections to unethical behavior.

se no moral dimension that is strong enough to push cooperation beyond a level reached with a Baseline VCM. Saliently framing an implicit moral duty of tax compliance leads to significantly increased contributions to the public good, but this effect is rather short-lived. This is mainly due to an insufficient number of good example setters in the experiment's early stages, resulting in the copying of others' noncompliance. Second, public disclosure of pictures as a nonpecuniary sanction generally significantly reduces free-riding and can stabilize cooperation over a longer period of time. However, we find qualitative differences in shaming: with no guidance toward the morality of tax avoidance, acting against the socially optimal outcome (full contributions) is perceived as less shame-worthy, compared with morally loaded tax avoidance.

With our experimental design, we clearly identify the shaming effect: in contrast to research on illegal behavior, disclosure is not confounded with elements of deterrence or strategic concerns resulting from the (moderate) risk of getting caught. Transferring our results to policy debates, the role of public exposure suggests that (the threat) of disclosing unwarranted behavior is an effective strategy for reducing tax avoidance and, more generally, promoting prosocial behavior. The communication of moral arguments also works (but is rather short-lived) and reinforces the shaming of tax avoiders.

In our experiment, shaming leaves subjects better off in terms of monetary payoffs. This does not imply, however, that the high social pressure it obviously induces is welfare-increasing in general (also see Dellavigna *et al.*, 2012 for a related point). Moreover, pillorying could be questionable outside the lab and would quickly collide with concerns about privacy and human rights, especially when it sets in at the slightest incidence of wrongdoing. Concerning real-world equivalents, however, the concept of (institutionalized) shaming has recently received heightened attention in the context of taxation (see the 'Related literature' section). Our results from the lab affirm the general (if transitory) efficacy of such measures. Whether the gains can outweigh the cost of the pillory obviously is a question that cannot be answered in a laboratory.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/bpp.2021.9>.

Acknowledgments. I thank Andreas Wagener and Kay Blaufus for their support and invaluable input. I thank Susan Steiner, Björn Jahnke, Sarah Ciaglia, and conference participants in Essen, Mannheim, Göttingen, Tampere and Freiburg for their inspiring comments and two anonymous referees for their helpful suggestions. I also thank Michael Milde for technical and organizational support.

Financial support. This study was funded by the German Research Council (DFG) in its Research and Training Group 1723 'Globalization and Development'. The DFG had no role in the design, execution, or analysis of this study.

References

- Alm, J., G. H. McClelland and W. D. Schulze (1992), 'Why do people pay taxes?' *Journal of Public Economics*, **48**: 21–38.
- Alm, J., M. Bernasconi, S. Laury, D. J. Lee and S. Wallace (2017), 'Culture, compliance, and confidentiality: Taxpayer behavior in the United States and Italy', *Journal of Economic Behavior & Organization*, **140**: 176–96.

- Andreoni, J. and D. Bernheim (2009), 'Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects', *Econometrica*, **77**: 1607–36.
- Andreoni, J. and R. Petrie (2004), 'Public goods experiments without confidentiality: A glimpse into fund-raising', *Journal of Public Economics*, **88**: 1605–23.
- Baldray, J. R. (1986), 'Tax evasion is not a gamble: A report on two experiments', *Economics Letters*, **22**: 333–5.
- Blaufus, K., J. Hundsdoerfer, M. Jacob and M. Suenwoldt (2016), 'Does legality matter? The case of tax avoidance and evasion', *Journal of Economic Behavior & Organization*, **127**: 182–206.
- Blaufus, K., J. Bob, P. E. Otto and N. Wolf (2017), 'The effect of tax privacy on tax compliance – An experimental investigation', *European Accounting Review*, **64**: 1–20.
- Blumenthal, M., C. Christian and J. Slemrod (2001), 'Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota', *National Tax Journal*, **54**: 125–38.
- Bø, E. E., J. Slemrod and T. O. Thoresen (2016), 'Taxes on the internet: Deterrence effects of public disclosure', *American Economic Journal: Economic Policy*, **7**: 36–62.
- Bochet, O., T. Page and L. Putterman (2006), 'Communication and punishment in voluntary contribution experiments', *Journal of Economic Behavior & Organization*, **60**: 11–26.
- Bock, O., I. Baetge and A. Nicklisch (2014), 'hroot: Hamburg registration and organization online tool', *European Economic Review*, **71**: 117–20.
- Bott, K. M., A. W. Cappelen, E. Sørensen and B. Tungodden (2017), 'You've got mail: A randomised field experiment on tax evasion', Discussion Paper SAM 10 2017. Bergen: Norwegian School of Economics, Department of Economics.
- Casal, S. and L. Mittone (2016), 'Social esteem versus social stigma: The role of anonymity in an income reporting game', *Journal of Economic Behavior & Organization*, **124**: 55–66.
- Coricelli, G., M. Joffily, C. Montmarquette and M. Villeval (2010), 'Cheating, emotions, and rationality: An experiment on tax evasion', *Experimental Economics*, **13**: 226–47.
- Dal Bó, E. and P. Dal Bó (2014), "'Do the right thing": The effects of moral suasion on cooperation', *Journal of Public Economics*, **117**: 28–38.
- Dellavigna, S., J. A. List and U. Malmendier (2012), 'Testing for altruism and social pressure in charitable giving', *The Quarterly Journal of Economics*, **127**: 1–56.
- Dimant, E. (2015), 'On Peer Effects: Behavioral Contagion of (Un)ethical Behavior and the Role of Social Identity', MPRA Paper No. 68732.
- Durham, Y., T. S. Manly and C. Ritesma (2014), 'The effect of income source, context, and income level on tax compliance decisions in a dynamic experiment', *Journal of Economic Psychology*, **40**: 220–33.
- Dwenger, N. and L. Treber (2018), 'Shaming for tax enforcement: Evidence from a new policy', Hohenheim Discussion Papers in Business, Economics and Social Sciences No. 21.
- Dwenger, N., H. Kleven, I. Rasaul and J. Rincke (2016), 'Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany', *American Economic Journal, Economic Policy*, **8**: 203–32.
- Fellner, G., R. Sausgruber and C. Traxler (2013), 'Testing enforcement strategies in the field: Threat, moral appeal and social information', *Journal of the European Economic Association*, **11**: 634–60.
- Fischbacher, U. (2007), 'Z-tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics*, **10**: 171–8.
- Fortin, B., G. Lacroix and M. Villeval (2007), 'Tax evasion and social interactions', *Journal of Public Economics*, **91**: 2089–112.
- Gino, F., S. Ayal and D. Ariely (2009), 'Contagion and differentiation in unethical behavior', *Psychological Science*, **20**: 393–8.
- Hallsworth, M., J. A. List, R. D. Metcalfe and I. Vlaev (2017), 'The behavioralist as tax collector: Using natural field experiments to enhance tax compliance', *Journal of Public Economics*, **148**: 14–31.
- Kimbrough, E. O. and A. Vostroknutov (2016), 'Norms make preferences social', *Journal of the European Economic Association*, **14**: 608–38.
- Lazear, E. P., U. Malmendier and R. A. Weber (2012), 'Sorting in experiments with application to social preferences', *American Economic Journal: Applied Econometrics*, **4**: 136–63.
- Luttmer, E. F. P. and M. Singhal (2014), 'Tax morale', *Journal of Economic Perspectives*, **28**: 149–68.
- Meiselman, B. S. (2018), 'Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment', *Journal of Public Economics*, **158**: 180–93.

- Noussair, C. and S. Tucker (2007), 'Public observability of decisions and voluntary contributions in a multi-period context', *Public Finance Review*, **35**: 176–98.
- Perez-Truglia, R. and U. Troiano (2015), *Shaming tax delinquents: Theory and evidence from a field experiment in the United States*, NBER Working Paper No. w21264. Cambridge, MA: National Bureau of Economic Research.
- Rege, M. and K. Telle (2004), 'The impact of social approval and framing on cooperation in public good situations', *Journal of Public Economics*, **88**: 1625–44.
- Samek, A. S. and R. M. Sheremeta (2014), 'Recognizing contributors: An experiment on public goods', *Experimental Economics*, **17**: 673–90.
- Tyran, J. R. and L. P. Feld (2006), 'Achieving compliance when legal sanctions are non-deterrent', *The Scandinavian Journal of Economics*, **108**: 135–56.
- Wang, Y., G. Wang, Q. Chen and L. Li (2017), 'Depletion, moral identity, and unethical behavior: Why people behave unethically after self-control exertion', *Consciousness and Cognition*, **56**: 188–98.
- Wartick, M. L., S. A. Madeo and C. C. Vines (1999), 'Reward dominance in tax-reporting experiments: The role of context', *Journal of the American Taxation Association*, **21**: 20–3.

Appendix

See [Tables A1](#) and [A2](#).

Table A1. Summary statistics for individual characteristics by treatment.

Treatment	Female (%)	Econ. Major (%)	Bachelor (%)	Employed (%)	Tax decl. (%)	Age	Income	Semester
Baseline	44.00	24.00	40.00	38.00	72.00	24.29	299.85	7.06
						(3.07)	(225.31)	(4.41)
Tax frame	52.50	20.00	50.00	27.50	65.00	24.30	366.78	7.78
						(2.71)	(261.30)	(3.37)
Moral frame	44.44	20.00	34.55	43.64	69.09	23.96	321.98	7.09
						(2.57)	(253.24)	(3.14)
Shaming	54.55	25.45	29.09	34.54	56.36	23.94	341.64	6.51
						(3.13)	(248.03)	(3.70)
Tax shaming	48.89	15.55	44.44	34.09	55.55	24.26	343.98	7.40
						(3.20)	(228.98)	(4.52)
Moral shaming	47.27	14.55	36.36	28.30	52.73	24.54	409.72	7.45
						(7.64)	(285.91)	(4.64)

Note: See [Table 2](#) for a description of the socioeconomic variables. Standard errors are in parentheses.

Table A2. Nonparametric Mann–Whitney *U*-tests for [Table 4](#).

Full compliance	10 rounds	7–9 rounds	3–6 rounds	0–2 rounds
Baseline				
Tax frame	–6***	5.5***	14***	–13.5***
Moral frame	–4.18***	3.45***	30.36***	–29.64***
Shaming	24.91***	25.27***	19.45***	–69.64***
Tax frame				
Moral frame	1.82***	–2.05	16.36***	–16.14***
Shaming	30.91***	19.77***	5.45***	–56.14***
Tax shaming	15.56***	12.5***	–2.22	–25.83***
Moral frame				
Shaming	29.09***	21.82***	–10.91***	–40***
Tax shaming	13.74***	14.55***	–18.58***	–9.69***
Moral shaming	36.36***	14.55***	–9.09***	–41.81***
Tax shaming				
Moral shaming	22.62***	0	9.49***	–32.12***

*** $p < 0.01$.

Notes: The numbers represent differences in percentage points between treatment and reference category (in bold letters). See [Table 4](#) for the respective frequencies of full compliance in the reference categories.