

# Tax Morale and the Role of Social Norms and Reciprocity – Evidence from a Randomized Survey Experiment\*

PHILIPP DOERRENBERG

ANDREAS PEICHL

September 11, 2020

## Abstract

We present the results of a survey (representative for the German population) with randomized components in the context of tax compliance. We first document the anatomy of tax morale and present novel evidence that previously found gender differences in tax morale are not driven by gender differences in risk preferences. The experimental component of our survey allows us to investigate the role of social norms and reciprocity for tax morale. One experimental treatment manipulates the social norm of tax compliance through the provision of information about the magnitude of the tax gap. The second treatment adds an appeal to reciprocity; it informed by how much education expenditures could be increased if the tax-gap-induced foregone revenue was available to the government. We find that the social-norm treatment reduces tax morale, while the reciprocity treatment has a positive effect on tax morale. This suggests that a potential backfire effect of social norms is outweighed if the consequences of violating the social norm are made salient in an appeal to reciprocity.

**JEL Classification:** H20, H32, H50, C93

**Keywords:** Tax compliance, Tax evasion, Intrinsic motivations, Tax morale, Social norms, Reciprocity, Survey Experiment

---

\***Doerrenberg:** University of Mannheim, CESifo, ZEW and IZA. Email: doerrenberg@uni-mannheim.de. **Peichl:** ifo Institute, University of Munich, CESifo, IZA and IHS. Email: peichl@econ.lmu.de. Eva Schoenwald and Felicitas von Dombos provided excellent research assistance. We thank Zareh Asatryan, Sebastian Blesse, Albrecht Bohne, Sebastian Camarero Garcia, Julie Cullen, Mathias Dolls, Denvil Duncan, Christoph Feldhaus, Roger Gordon, Paul Hufe, Ilyana Kuziemko, Armando Meier, Daniel Mueller, Carina Neisser, Joerg Paetzold, Emmanuel Saez, Jan Schmitz, Joel Slemrod, Holger Stichnoth, Christian Traxler and Nils Wehrhoefer as well as various seminar participants for helpful comments and suggestions. This paper uses data from the German Internet Panel wave 14 (DOI: 10.4232/1.12620; Blom et al. 2016). A study description can be found in Blom et al. (2015). The German Internet Panel (GIP) is funded by the German Research Foundation (DFG) through the Collaborative Research Center 884 "Political Economy of Reforms" (SFB 884).

# 1 Introduction

It is now widely acknowledged that the decision to evade taxes is not only driven by extrinsic, pecuniary factors (such as tax rates, penalties, audit probabilities and enforcement), but also by intrinsic, non-pecuniary motives.<sup>1</sup> Following Luttmer and Singhal (2014), we use the term *tax morale* as an umbrella term for such intrinsic, non-pecuniary tax-compliance motives.<sup>2</sup> In light of evidence on the role of tax morale for tax compliance (see footnote 1 and below), it has been argued that tax morale has the potential to ensure tax compliance in situations where third-party reporting and the capacity to control and audit taxpayers are limited (Luttmer and Singhal 2014; Bott et al. 2019). An important question then is, which factors shape tax morale. In this paper, we add to this question and study tax morale in a representative survey with randomized components. The survey allows us to shed new light on the overall anatomy of tax morale and to explore the causal link between tax morale and two potential drivers: social norms and reciprocity.

Social norms and reciprocity are often believed to be key determinants of tax morale (Luttmer and Singhal 2014).<sup>3</sup> Reciprocity in this context means that the motivation to comply may depend on the (perceived) quality of government services and public goods which citizens receive in return for their tax payment (see Besley 2020 for a model that formalizes the reciprocal relationship between citizens and the government in the context of taxation. This is also related to the concept of conditional cooperation, e.g. Frey and Meier 2004). Social norms of tax-compliance behavior particularly depend on the perception about the prevalence of tax evasion in society. Taxpayers might be more willing to evade if (they have the impression that) evasion is very common, and they might be more compliant if (they believe that) most other taxpayers pay their taxes honestly. The

---

<sup>1</sup>For example, Dwenger et al. (2016) provide evidence that a significant fraction of individuals comply with taxes even in the absence of any deterrence (i.e., in the absence of any penalties or audits). DeBacker et al. (2015) study a situation in which taxpayers differ in their level of intrinsic motivation but operate in the same deterrence environment. They provide evidence that taxpayers with lower intrinsic motivations indeed are less compliant. Frimmel et al. (2018) provide evidence of an intergenerational causal effect in tax-evasion behavior (i.e., parent tax evasion affects children tax evasion), implying that non-pecuniary factors (such as family ties) matter for evasion. Dulleck et al. (2016) provide neural evidence for the importance of non-pecuniary motives in tax-compliance decisions. Laboratory experiments also show that non-pecuniary motives matter for tax compliance (Alm and Malezieux 2020). The US Internal Revenue Service (IRS) mentions 'socio-political' factors as primary drivers of tax compliance (IRS 2007). Luttmer and Singhal (2014) provide a survey and summarize the role of non-pecuniary motives and intrinsic motivations on actual compliance in detail.

<sup>2</sup>Dwenger et al. (2016) use the term *intrinsic motivations for tax compliance* while other papers use *tax ethics* or *tax honesty* to describe what we label tax morale.

<sup>3</sup>Luttmer and Singhal (2014) provide a typology of tax-morale mechanisms in which they classify social norms and reciprocity to be among its key drivers. The definitions of social norms and reciprocity in Luttmer and Singhal (2014) are congruent with the following definitions which we use throughout our paper. Bott et al. (2019), for example, also have a focus on reciprocity and social norms in their tax-morale study. See below for more examples that underline the important role of social norms and reciprocity for tax morale.

randomized components of our survey are designed to study whether social norms and reciprocity have a causal effect on tax morale.

Our survey is embedded in the German Internet Panel (GIP), a representative online survey in Germany. We included a tax-morale question in the GIP that is similar to the frequently used tax-morale question in the World Values Survey (see, e.g., Slemrod 2003, Alm and Torgler 2006).<sup>4</sup> The rich nature of the GIP allows us to study the association between tax morale and a wide set of factors (such as demographics and preferences). Before responding to the tax-morale question, participants were randomly assigned to three experimental groups (with an augmented treatment structure). First, the question is preceded only by a general note that issues of tax evasion are often discussed in the media. We label this group the 'control group'. Second, in addition to this general note, participants are informed about scientific estimates of the tax gap in industrialized countries.<sup>5</sup> This experimental variation intends to manipulate the social norm of tax evasion by providing information about the level of tax evasion in the population. We label this group the 'social-norm group'.

Third, in addition to the information in the social-norm group, participants are informed by how much the government expenses for education in Germany could be increased if the foregone revenue that is due to the tax gap was spent on education. This variation adds a reciprocity component as it increases awareness and salience about the relationship between evaded taxes and government expenditures and services. It refers to the specific example of education expenditures, which are likely to be perceived as beneficial for society in general and maybe also for an individual in particular. Hence, by adding the potentially bad consequences of non-compliance, this treatment confronts survey participants with the social externalities of tax evasion. We label this group the 'reciprocity group'. None of the information provided in the treatments were deceptive (see Section 2 for a validation and justification of information and numbers referenced in the experimental treatments).

Our analysis starts with an exploration of the anatomy of tax morale. We confirm earlier findings on the (correlational) link between gender and age on the one hand side and tax morale on the other hand side: women tend to have higher tax morale and tax morale increases with age. The novelty of our paper is that we have measures of risk aversion and patience in our survey data. We show that neither gender nor age effects

---

<sup>4</sup>The question asks participants to which extent they find tax evasion justifiable in a hypothetical situation with an 'easy opportunity' to evade. It does not ask participants about their actual tax-compliance behavior. There is suggestive evidence indicating that the replies to the corresponding question in the WVS are indeed linked with actual levels of tax evasion (Halla 2012).

<sup>5</sup>The tax gap is a common measure for the extent of tax evasion (Slemrod 2007). It is defined as the share of outstanding taxes relative to actual (paid plus unpaid) tax liability. In order to make the treatment message comprehensible for a general audience, we do not actually use the word 'tax gap'. See section 2 for the exact wording of the treatment messages.

are driven by risk or patience. In addition, we confirm the intuitive expectation that risk aversion and tax morale are positively correlated. Also in line with intuition, as well as corresponding with recent results from the literature on attitudes towards redistribution (Alesina et al. 2018), we find that participants with right-wing political attitudes have significantly lower tax morale.

The randomized survey experiment exhibits the following main results. First, manipulating the social norm through information about the general extent of tax evasion has a negative effect on tax morale, relative to the control group (though this effect sometimes lacks statistical precision). This social-norm effect is in line with literature in other contexts (see below) and is consistent with the notion that social norms can backfire if they reveal that a certain behavior is regrettably frequent. Second, if an appeal to reciprocity is added to the social-norm information, tax morale becomes significantly larger (i.e., the effect of reciprocity relative to the social-norms treatment is significantly positive).<sup>6</sup> The treatment reminds participants that beneficial government services can only be provided *in return* for compliance among taxpayers. Our findings suggest that such an appeal to reciprocity works and that aspects of reciprocity matter for tax morale. In addition, it indicates that a backfire effect of social norms is outweighed if the consequences of the social norm are made salient.

The experimental treatment effects are in the range of 2-3 percentage points. Given that only 11% of participants find tax evasion acceptable, the effects are economically meaningful. Moreover, the magnitude of the effects should also be considered in light of the fact that tax morale is usually seen to be a fairly inelastic parameter which is shaped over a lifetime through experiences as a taxpayer, perceptions of and attitudes towards the government as well as culture (Luttmer and Singhal 2014). Furthermore, the interpretation of the size of the effects has to take into consideration that our experimental manipulation consisted of only one or two additional sentences and was therefore fairly minor.

We relate to several strands of the literature. First, we add to the general literature on tax evasion, and in particular to the work on intrinsic motivations for tax compliance (in addition to the recent survey by Luttmer and Singhal 2014, the overview article by Andreoni et al. 1998 also highlights the importance of non-pecuniary motives for compliance).<sup>7</sup> As already discussed in footnote 1 above, there is evidence that tax morale

---

<sup>6</sup>Note that we have an augmented treatment structure; this implies that the reciprocity treatment includes the information from the social-norm treatment *plus* the reciprocity component. To identify the effect of reciprocity, we thus need to compare the reciprocity treatment relative to the social-norm treatment. Comparing the reciprocity group to the control group does not identify the reciprocity effect because this comparison includes the effect of the social-norms treatment *and* the effect of the reciprocity treatment.

<sup>7</sup>The recent meta-analysis paper by Alm and Malezieux (2020) provides an overview of the rich tax-evasion literature using laboratory experiments, many of which also study non-pecuniary aspects and

matters and translates into actual tax-paying behavior. Such findings motivate studies on tax morale and make them relevant.<sup>8</sup> An earlier strand of tax-morale literature mainly uses data from the WVS to study its correlational determinants (see e.g., Torgler 2006). These papers find that tax morale is correlated with variables such as gender or age. We confirm these findings and add additional evidence on the anatomy of tax morale, for example with respect to personal characteristics such as risk aversion and patience. A particular contribution is our finding that neither gender nor age effects are driven by risk preferences or patience. Our paper joins two recent other studies that use survey experiments in the context of tax morale; these papers focus on the role of government performance (Ortega et al. 2016, see below for more) as well as anti-corruption policies and giving voice regarding expenditure preferences (Sjoberg et al. 2019).

Second, we relate to literature on the tax-compliance effects of social norms (compliance behavior of others) and reciprocity (highlighting the role of tax payments for government provided services and goods), which both belong to the typology of the main tax-morale ingredients that is put forward by Luttmer and Singhal (2014). Survey correlations and experimental evidence from the laboratory support the significance of reciprocity for compliance (e.g., Alm and Jackson 1993; Scholz and Lubell 1998; Frey and Torgler 2007; Cummings et al. 2009; Alm 2012; Lambertson et al. 2014). Two papers using observational data and natural experiments also suggest that government policy/approval affects compliance and, thus, that reciprocity may matter: Cullen et al. (2018) show that government approval has a positive effect on tax evasion, and Besley et al. (2015) document that the introduction of a tax that was widely perceived to be unfair increased evasion.<sup>9</sup> Using a survey with experimental components in different Latin American countries, Ortega et al. (2016) find that information about government achievements has a positive effect on the willingness to pay taxes; this finding is also indicative of a positive reciprocity effect. However, evidence from randomized field interventions on the role of reciprocity on tax compliance is somewhat mixed. While studies such as Blumenthal et al. (2001), Dwenger et al. (2016), Castro and Scartascini (2015) and Bergolo et al. (2017) do *not* find any significant effects of highlighting the services of the state/government, Bott et al. (2019) and Hallsworth et al. (2017) provide evidence for a positive reciprocity effect on compliance.

Evidence on the role of social norms for compliance is also mixed. Several field-

---

determinants of tax evasion (the most relevant laboratory experiments in our context are referenced below).

<sup>8</sup>Luttmer and Singhal (2014, page 151) also ”*argue that tax morale is indeed an important component of tax compliance decisions, though [they] view enforcement as the primary driver of compliance.*”

<sup>9</sup>Falkinger (1988) presents a theoretical model in which taxpayers value the shares of public goods that they receive. Congdon et al. (2009) stress in their overview article that it likely matters for tax behavior what people believe their taxes are used for. Additional empirical work is Cebula (2013) who shows a positive relation between IRS measures of aggregate tax evasion and government satisfaction.

experimental papers do not find any compliance effects of communicating that the majority of taxpayers is compliant with the law (e.g., Blumenthal et al. 2001; Fellner et al. 2013; Dwenger et al. 2016; Castro and Scartascini 2015; Perez-Truglia and Troiano 2015). In contrast to these null findings, other studies do find effects of social norms on compliance behavior. Hallsworth et al. (2017) document in a randomized setting that telling taxpayers that '9 out of 10 people (in the UK) pay their taxes on time' increases punctual payment of tax debt. Bott et al. (2019) find a positive compliance effect of a letter stating that the majority of Norwegian taxpayers report income correctly. Paetzold and Winner (2016) show that taxpayers evade more taxes after they change jobs to a firm where evasion is more common than in their previous firm – this also suggests that social norms and the behavior of others matter for tax compliance.<sup>10</sup>

In light of these mixed findings in the compliance literature, our paper adds new evidence on the role of social norms and reciprocity for tax compliance. While several recent studies use randomized variation in the field (Kleven et al. 2011; Pomeranz 2015) and look at reported tax bases as outcome variables, we implement a randomized survey experiment with a focus on survey-reported tax morale. One particular advantage of conducting a survey experiment is that we are able to study social norms and reciprocity within the same design and compare their relative importance for tax morale. In contrast to field experiments, the survey experiment further allows us to examine attitudes towards evasion (rather than actual behavior), which might be more sensitive to small interventions than actual tax payments that involve large stakes. As a result, the survey experiment could be used to identify possible effects of social norms and reciprocity in the context of compliance which can hardly be identified in a field experiment. This relates to the argument by Luttmer and Singhal (2014) that null findings of moral appeals in evasion field experiments should be attributed to the lack of power of the interventions rather than suggesting that moral appeals do not matter. The small interventions in our survey experiment potentially have enough power to manipulate attitudes towards tax evasion (while similar interventions in the field are not sufficiently powerful to shift actual behavior).<sup>11</sup>

Third, our paper relates to the extensive literature on the role of social norms and

---

<sup>10</sup>In line with this empirical evidence on the existence of social norms, Traxler (2010) incorporates social norms into the standard Allingham and Sandmo (1972)-model and models them as depending on the tax-compliance behavior of other citizens (whereby more evaders in the society increase the individual willingness to evade). Fortin et al. (2007), Chetty et al. (2013) and Bohne and Nimczik (2018) are further examples from tax-related contexts showing that the behavior of other people affects individual behavior.

<sup>11</sup>Our survey-based approach to study attitudes towards tax evasion is also related to the literature on attitudes towards redistribution which mainly uses survey questions to identify the drivers of redistributive attitudes (examples for this literature include Luttmer 2001; Corneo and Gruener 2002; Fong and Luttmer 2011). As we do in our study, this literature recently also implemented randomized survey experiments to shed light on the drivers and elasticity of attitudes with respect to information (Cruces et al. 2013; Kuziemko et al. 2015; Alesina et al. 2018). For a recent overview of the growing literature using survey experiments see Haaland et al. (2020).

reciprocity in other (non-tax) contexts. For example, the literature on public goods – where contributing to the public good can be interpreted as the equivalent choice to paying taxes honestly – has shown that people contribute more to the public good the more others contribute (e.g., Weimann 1994, Keser and Van Winden 2000) and the more they expect in return for contributing to the public good (e.g. Zelmer 2003). That is, social norms and reciprocity seem to matter for public-good provision. The literature strands on charitable giving and pro-environmental behavior also show that social norms and reciprocity matter and that they increase the likelihood of choosing the desired ‘more moral’ options such as higher donations or saving more energy (e.g. Andreoni and Scholz 1998; List and Lucking-Reiley 2002; Frey and Meier 2004; Allcott 2011). Our paper adds evidence that the effects of social norms and reciprocity, which are found in different contexts of moral behavior, also translate to moral behavior in the context of tax compliance.

The rest of the paper is organized as follows. Section 2 describes the survey and experimental variations. The results of the randomized survey experiment are presented and discussed in Section 4. Section 3 presents the results with respect to the anatomy of tax morale. Section 5 concludes.

## 2 Survey and experimental treatments

**The survey.** We collected survey data through the German Internet Panel (GIP). The GIP is a longitudinal survey that is operated and administered at the University of Mannheim in Germany.<sup>12</sup> The main purpose of the panel survey is to collect “data on individual attitudes and preferences relevant in political and economic decision making processes”. GIP data are collected online on a bi-monthly basis. The survey is representative for the German population aged 16 to 75.<sup>13</sup>

The survey includes repeated questions (included in every wave) as well as questions only included in single waves. We included the question on tax-compliance attitudes that is at the center of this paper in wave 14 (the relevant question is numbered CF14015, see Blom et al. 2016). This wave went to the field in November 2014 and included 3,575 participants. The data were released in 2016. For our analysis, we use wave 14 and

---

<sup>12</sup>To be more precise, the survey is based at the “Collaborative Research Center 884 on Political Economy of Reforms”, which is funded by the German Science Foundation (*Deutsche Forschungsgemeinschaft*, SFB 884). See <http://reforms.uni-mannheim.de/> for background information on the research center. Also see the general survey description in Blom et al. (2015) and at [http://reforms.uni-mannheim.de/internet\\_panel/home/](http://reforms.uni-mannheim.de/internet_panel/home/). Examples of GIP-based papers include Kerschbamer and Müller (2017), Müller and Renes (2017), Dolls and Wehrhoefer (2018) and Engelmann et al. (2018).

<sup>13</sup>Recruitment was conducted offline with face-to-face interviews, during which respondents were invited to the online panel. To ensure the representativeness of the sample, the GIP includes respondents without prior computer or Internet access by providing them with the necessary equipment and training.

complement it with demographic information surveyed in previous waves.

**Measurement of tax morale.** Our measure of tax morale is based on the following GIP question:

*How justifiable do you think it is to evade taxes if an easy opportunity to do so presents itself?*

Survey participants can reply to this question on a 6-point scale. The reply categories were: *very justifiable*, *justifiable*, *rather justifiable*, *rather not justifiable*, *not justifiable* and *not at all justifiable*.<sup>14</sup> The first two categories of the tax-morale variable, *very justifiable* and *justifiable*, were only selected by 4.65% of all survey participants. The corresponding variable in the WVS (see below) is distributed similarly with only very few survey participants choosing that tax evasion is (very) justifiable. The literature using the WVS therefore usually uses recoded versions of the tax-morale variable (e.g., Torgler 2006; Alm and Torgler 2006; Halla 2012). In order to account for the distribution of the original survey question, we follow this approach and use a binary version of the variable as the main outcome variable in our empirical analysis. The recoded dummy variable which we use takes value "1" for respondents who find tax evasion *not at all justifiable*, *not justifiable* or *rather not justifiable* and it takes value "0" for respondents who find tax evasion *very justifiable*, *justifiable* or *rather justifiable*. That is, we create a dummy variable which indicates if a respondent has high tax morale (evasion is more or less not justifiable) vs. low tax morale (evasion is more or less justifiable). This dummy variable allows for an intuitive and simple interpretation of the results, and it increases statistical power. We use the 6-point scale in a robustness check below.

The question is an adapted version of the commonly used WVS question.<sup>15</sup> This question creates a hypothetical situation in which taxpayers have an 'easy opportunity' to evade and it does not ask participants about their actual tax-compliance behavior. The hypothetical character of the question ensures that participants will not fear any consequences from indicating that they find evasion acceptable (as they maybe would if they were asked for actual evasion behavior) and therefore intends to trigger reliable an-

---

<sup>14</sup>The original question in German was: *Fuer wie vertretbar halten Sie es, Steuern zu hinterziehen, wenn sich dafuer eine einfache Moeglichkeit ergibt?* The original reply categories were: *fuer sehr vertetbar*, *fuer vertetbar*, *fuer eher vertetbar*, *fuer eher nicht vertetbar*, *fuer nicht vertetbar* and *fuer ueberhaupt nicht vertetbar*. The question and answers were designed by the administrators of the survey who have an extensive and long-standing expertise in survey methodology building on a similar question in the WVS (see below).

<sup>15</sup>The WVS question was for example used in Slemrod (2003), Alm and Torgler (2006), Richardson (2006), Torgler (2006) and Halla (2012). It reads: *Please tell me for the following statement whether you think it can always be justified, never be justified, or something in between: 'Cheating on taxes if you have the chance'.*

swers which reflect the true intrinsic motivation to pay taxes.<sup>16</sup> Indeed, there is empirical work suggesting that replies to the equivalent question in the WVS are associated with actual levels of tax evasion and the shadow economy (Torgler and Schneider 2009; Halla 2012).

**Randomized survey experiment.** Before replying to this survey question on tax compliance, all participants were randomly assigned to three different groups in a between-subjects design; 'control group', 'social-norm group' and 'reciprocity group'. Screenshots of the three experimental conditions are displayed in Appendix Figures 2, 3 and 4. We have an augmented treatment structure where we subsequently add information.<sup>17</sup> That is, participants in the social-norm group receive the same information as participants in the control group plus additional information, and participants in the reciprocity group receive the same information as participants in the social-norm group plus additional information.

In the **control group**, the survey question was only preceded by a short opener stating that cases of tax evasion are frequently discussed in the media. This opener served the purpose of a short introduction to the question and a brief motivation for its relevance. It also ensured that the tax compliance question does not come out of the blue. Almost all questions in the GIP are preceded by a short comparable opener. 1,178 out of 3,532 participants were assigned to this control group.

In the **social-norm group**, the opening sentence in the control group was complemented with a statement about the prevalence of tax evasion: *Scientific studies estimate that in industrialized countries approximately 10% of all taxes which the government is entitled to are being evaded.* By providing reliable information about the commonness of tax evasion, these information are intended to manipulate the social norm of tax evasion. Providing subjects with a number describing the commonness of a certain type of behavior is the usual approach in the experimental literature for manipulating social norms (e.g., Frey and Meier 2004).<sup>18</sup> This social-norm treatment relates to the typology of tax-morale mechanisms in Luttmer and Singhal (2014) who classify the "views or behaviors

---

<sup>16</sup>In light of the formulation of the question asking for the 'justifiability' of evasion, it seems possible that some participants think that evasion can be 'justified' (for whatever reason) but still is not something that they would want to be engaged in. We acknowledge that the question does not necessarily reveal anything about an actual private action. However, even a shift in 'justifiability' that does not come with a shift in actual private behavior is potentially interesting and reveals something about attitudes towards evasion.

<sup>17</sup>A 2x2 between design probably would have lacked power.

<sup>18</sup> One might also refer to this experimental manipulation as a 'social information treatment', 'conditional cooperation treatment' or a 'descriptive norm treatment'. We use the wording 'social norm' in line with previous literature that (randomly) provides information about the behavior of others to manipulate 'social norms' (e.g., Allcott 2011; Hallsworth et al. 2017). Our way of manipulating social norms in the context of tax-paying behavior is comparable to one of the treatments in Hallsworth et al. (2017) in which they communicate taxpayers that 'Nine out of ten people pay their tax on time'.

of other individuals” as one important mechanism in this typology. The strength of the social-norm manipulation depends on participants’ priors about the extent of tax evasion; the larger the difference between the prior and the number presented in the treatment, the stronger is the shift in the social norm. Unfortunately, it was not feasible to ask participants about their initial priors. Tax enforcement in Germany has a solid reputation, suggesting that the average participants’ initial believe was that less than 10% of taxes are evaded – but this is only speculation.

The most reliable information about the magnitude of the tax gap come from random audit programs. These are rare and their results oftentimes not published. Unfortunately, Germany does not conduct such randomized audit programs. The overview article by Slemrod (2007) summarizes the available information about tax gaps estimated from such randomized audit programs. The 2001 net tax gap in the US was estimated to be 16.3% of estimated actual (paid plus unpaid) tax liability. A European country with a random audit program is Sweden, where the 1997 tax gap was estimated to be 9%. An official document from the UK speculates that the UK tax gap is of similar magnitude to that of Sweden and the United States (reported in Slemrod 2007). In light of these information, we opted for providing the information that the tax gap in industrialized countries is *approximately* 10%. Given the magnitudes of 16.3% for the US and 9% for Sweden, 10% appears to be rather conservative, ensuring that we do *not* provide information that are too extreme or deceptive in any way. 1,177 out of 3,532 participants were assigned to this social-norm group.<sup>19</sup>

Participants in the **reciprocity group** received the same information as participants in the social-norm group. In addition, they faced the following statement: *With these foregone earnings, the German government could raise its expenditures for education by about 50 percent.* This treatment highlighted that the extent of tax evasion has implications for government budget and makes it salient to participants that tax evasion

---

<sup>19</sup>There are two potential caveats to our formulation of the social-norm treatment. First, while a 10% tax gap might surprise the average citizen, individuals who are at the margin to saying tax evasion is justifiable are probably those who have a particular negative view of tax enforcement. However, if these individuals believed the tax gap to be larger compared to the average respondent, they should respond less to our intervention (or even in the opposite direction). As a result, our treatment estimates would be attenuated. Second, among those that are surprised by the provided tax gap magnitude, it could be the case that some conclude from the updated information that tax evasion is less prevalent than they previously thought because the risk of getting caught is much smaller than what they believed it to be. As a consequence, these individuals might find tax evasion more justifiable in response to the treatment information only for reasons of private income maximization. While it is impossible to rule out this potential channel, note that it would have been difficult, if not impossible, to give information about the compliance behavior of others, while holding the perceived detection probability constant. We therefore decided to comply with the usual approach in the previous literature on social-norm interventions (see above, in particular footnote 18). Moreover, we are interested in potential differences between the reciprocity treatment (see below) and the social-norms treatment. Such differences cannot be explained by this argument, because participants were informed about the tax gap in the same way both in the social-norm treatment as well as the reciprocity treatment.

potentially has immediate consequences for the provision of public goods through the government; the treatment hence reminds participants that beneficial government services can only be provided *in return* for compliant tax payers.<sup>20</sup> We chose education expenditures as an example because this policy field is widely acknowledged to be important and to create value; most people in Germany likely agree that higher education expenses are better than lower education expenses (as for example reflected in the discussions in basically all election campaigns). A large survey in Germany with more than 400.000 respondents (Zukunft durch Bildung 2011) shows that the vast majority finds education to be 'extraordinary important' and that 73% of the participants are willing to pay higher taxes for the improvement of education in Germany.

The treatment reminds participants that more compliance could yield better policy in return, and thus appeals to the reciprocity of participants.<sup>21</sup> The treatment indirectly reminds participants that the 'implicit contract' between the government and the citizens (Feld and Frey 2007) – i.e., people pay taxes and receive government services in return – is threatened through tax evasion. This treatment also relates to the typology of Luttmer and Singhal (2014) which classifies reciprocity – defined as an additional utility term for paying taxes honestly that depends in some way on the individual's relationship to the state – to be one of the key ingredients of tax morale. It was again not possible to survey participants' priors or whether they were (positively or negatively) surprised by the treatment information. We speculate that most participants had not realized that government services are potentially damaged through tax evasion to such a large extent, suggesting that tax morale increases in response to the information treatment.

The information that government expenses for education could be increased by *about 50%* if the tax-gap induced foregone earning were to be spent on education were calculated as follows. Total tax revenues in 2013 amounted to about 620 billion Euro.<sup>22</sup> A tax gap of 10% then implies that the foregone revenues due to the tax gap amounts to about 69 billion Euro (tax gap =  $0.1 = 69/(620 + 69)$ ). According to the Federal

---

<sup>20</sup>This type of reciprocity statement is comparable to one of the treatments in the field experiment by Bergolo et al. (2017). Subjects in their respective treatment are also told by how much government expenses for particular 'good' policies could increase if there was less evasion. As in Bergolo et al. (2017), we do not intent to give the impression in our reciprocity treatment that all extra revenue from less evasion would indeed spent on education. The treatment shall simply remind participants that evading has significant implications for government spending using education expenses as an illustrative example.

<sup>21</sup>Fehr and Gaechter (2000, page 159) define reciprocity in their survey article as follows: "*Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model.*" Our reciprocity treatment stresses the mutual dependence of tax compliance and government services and reminds people that they should be "nice and cooperative" (i.e., pay taxes) because the government does "friendly actions" (i.e., provide education) in response to cooperation and tax honesty.

<sup>22</sup>Source: [https://www.destatis.de/DE/Publikationen/Thematisch/FinanzenSteuern/Steuern/Steuerhaushalt/SteuerhaushaltJ2140400137004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/FinanzenSteuern/Steuern/Steuerhaushalt/SteuerhaushaltJ2140400137004.pdf?__blob=publicationFile)

Statistical Office, the expenses for education in Germany in 2013 stood at 116 billion Euro.<sup>23</sup> These numbers then imply that education expenses would have increased by 59% (= 69/116) if all foregone revenues (69 billion) were to be spent on education. In order to provide a conservative estimate and again insure ourselves against any type of deception, we chose to give the information that education expenses could increase by *about* 50%. 1177 out of 3532 overall participants were assigned to this reciprocity group.

A potential concern with most (survey) experiments is that experimenter effects could drive some of the findings. Two recent papers by de Quidt et al. (2018) and Mummolo and Peterson (2019) study experimenter effects in survey experiments explicitly and provide evidence that experimenter demand apparently is not much of a concern in survey experiments.<sup>24</sup> While we find these results very reassuring, we additionally identify four arguments in the context of our specific study for why experimenter demand effect may not be a major concern. First, our experiment is a between-subject design where every respondent is in either one of the three experimental groups. That is, the experimental intervention is not made salient and respondents are not aware that other respondents receive questions with different contents. As a result, they are not induced to think that they are ought to give a particular answer. Second, the GIP surveys many different general attitudes and opinions and does not give participants the impression that there are correct or false replies. Third, even if there were experimenter effects, the comparison of the social-norms treatment and the reciprocity treatment would still be valid because it is very unlikely that the social-norms treatment message induced different experimenter effects than the reciprocity treatment. In other words, a potential experimenter effect is likely held constant between the experimental groups. Fourth, participants fill out the survey without an interviewer that is physically present.

By nature, the effects of our experimental interventions are measured in the short run and it remains unclear how persistent they are. However, the purpose of our survey experiment is not to investigate whether these interventions can manipulate behavior persistently. We instead aim to study if attitudes respond to small information treatments. It is intended on purpose that the information interventions are salient, fresh on participants' minds and well remembered when the participants respond to the question. If we were not certain that participants indeed remembered the information in our interventions, it would be difficult to identify the effect of these information on attitudes. In this regard, the purpose of our survey experiment is different than the purpose of field

---

<sup>23</sup>Source: [https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2014/02/PD14\\_066\\_217.html](https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2014/02/PD14_066_217.html)

<sup>24</sup>Mummolo and Peterson (2019) run online survey experiments with more than 12,000 participants and randomly assign information about experimenter intent. They find that providing these information does not affect treatment effects; even financial incentives to respond in line with experimenters' intent did not trigger any demand effects. The findings are consistent with the results of de Quidt et al. (2018). They use a similar approach in online experiments and find that experimenter demand effects are 'small'.

studies that investigate the effects of information treatments on actual behavior (also see the discussion in the introduction with respect to the complementarity of field and survey studies).

**Variable description and summary statistics.** Table 1 provides an overview of all variables which we use throughout the paper (including measurement and category information). Table 2 presents summary statistics for these variables. The table shows that mean tax morale across all participants is at 0.89 (with standard deviation 0.32), meaning that 89% of participants indicate that tax evasion is not at all/not/rather not justifiable. Attrition is not an issue with the tax morale question; only about 1% of respondents have a missing value for this question.

With regard to the other variables, Table 2 further shows that we have a balanced share of men and women in the sample, 58% of all participants are married and the average household of participants has 2.50 members. Age is only measured in categories (see table 1) and we see a roughly even distribution across the age categories (18% of participants are younger than 30 and 24% are older than 60 years old). The share of retired participants is 16%, most participants are in income category 2 (40% with net household income between 1,500 and 3,000 Euro) and education category 3 (52% with high school with university qualification or apprenticeship), and their political preferences are mostly conservative or moderate left. In line with low unemployment rates in Germany, only 3% of people in our sample are unemployed. 20% of our survey participants live in East Germany.

The GIP survey contains questions on risk attitudes and patience. The according questions ask participants about their general willingness to take risks and their general level of patience. This raises the natural question of whether these self-reported survey questions are a reliable predictor of actual behavior. Evidence in this direction for the risk variable is provided by Dohmen et al. (2011) who compare survey questions on self-reported risk with actual risk-taking behavior using a representative population of the adult population in Germany. They elicit actual risk-taking behavior through an incentivized real-stakes lottery experiment and their self-reported survey measure of risk is very similar to the risk question in our GIP survey. Their results provide strong evidence that the responses to the survey risk question are a strong predictor of actual risk behavior, even controlling for a large number of observables.<sup>25</sup> Vischer et al. (2013) validate a self-reported patience question against actual behavior by. They use a similar

---

<sup>25</sup>The authors' own conclusion is that these "findings document that a simple, qualitative survey measure can generate a meaningful measure of risk attitudes, which maps into actual choices in lotteries with real monetary consequences. This is important because it suggests that surveys can collect information on risk attitudes using instruments that are easy to use and relatively cheap to administer, and yet deliver a behaviorally valid measure of risk attitudes" (Dohmen et al. 2011, page 524). The paper also studies if particular risky activities, such as holding stocks, being self-employed or smoking, are correlated with the survey question on risk attitudes. The results show that the general self-reported risk question is a good all-round explanatory variable for predicting all behaviors.

approach as Dohmen et al. (2011) and compare survey responses to actual behavior in incentive-comparable inter-temporal choice experiments. They find that the survey responses indeed predict actual behavior. In addition, the Global Preference Survey (GPS), a large cross-country survey to measure economic preferences, includes survey questions on risk aversion and time discounting, which are carefully validated using incentivized choice experiments (Falk et al. 2016; Falk et al. 2018). The questions in the GPS are comparable to the GIP questions that we use to measure risk and patience – this is further support that our survey questions are reliable measures for actual preferences with respect to risk and patience.

We take the findings of these studies as reliable evidence that our survey responses for risk attitudes and patience are appropriate measures for actual risk and patience behavior. With regard to summary statistics for these two variables, Table 2 shows that the average level of risk aversion is at 3.67 and average patience is at 3.48, both measured on a 5-point scale.

**Randomization checks.** Following the strategy in Alesina et al. (2018), we test balance across experimental groups as follows: For each covariate, we run three OLS regressions of the form  $y_i = \beta Covariate_i + \epsilon_i$ , where *Covariate* is the respective covariate that we test. The three dependent variables for which we run the regressions are dummies indicating the treatment groups – control, social norm, reciprocity. We test this for 9 covariates which measure pre-defined demographics. As a result of this procedure, we have the results of 27 OLS regressions (one regression for each combination of 9 covariates and 3 outcome dummies).

Appendix Table 6 shows the p-values for these 27 regressions (robust standard errors). Overall, randomization worked very well. We present 60 coefficients in this regression table and find only five coefficients to be statistically significant. Two out of these five coefficients belong to the unemployment variable: we do see a significant effect of the unemployment variable on the probability of belonging to the control and social-norm group. As the summary statistics showed, the share of unemployed people in the sample is only 3% and we only have a few unemployed individuals in each of the three treatment groups (19 unemployed in control group, 41 in social-norm group and 32 in reciprocity group). This might explain a potentially unlucky randomization with respect to this variable and also implies that this variable is not a big concern.

In general, having a few variables that can explain treatment status is not unusual and often the result of a true randomization process. For example, around 5% of all presented randomization statistics are significant in Kleven et al. (2011, section 6.1). This is in a comparable range to the share of coefficients that are statistically significant in our set-up. To circumvent any concerns with regard to randomization as good as possible,

we show regression results with and without conditioning on covariates (including the employment status).

### 3 The anatomy of tax morale

Table 3 shows the estimates of a simple OLS regression of our outcome variable – tax morale – on different variables included in the survey.<sup>26</sup> These estimates are conditional correlations and should not be given a causal interpretation. However, they can shed light on the drivers of tax morale, thereby complementing other studies based on field experiments or tax-return data (which do not have information on many variables) and adding to the large survey literature based on the WVS. In addition to using another sample than the WVS, our survey has the advantage that we have a more precise measure of income as well as two variables which are likely to matter for compliance that are not included in the WVS: patience and risk aversion. The outcome variable in the regressions is a binary variable that takes value "1" if evasion is not justifiable and value "0" if evasion is justifiable (see above). All variables are measured and coded as explained in Table 1.

Specification (I) includes basic demographic variables as explanatory variables. These include gender, age, marital status, employment status, retirement status and education level. We then subsequently add further variables to the regression.<sup>27</sup> Specification (II) adds a net-household-income measure, Specification (III) adds two variables which reflect the character of a participant – risk aversion and patience –, and Specification (IV) adds political preferences. Specification (V) adds a categorical variable indicating the treatment group from the randomized survey.

In accordance with most other studies,<sup>28</sup> we find that women have higher tax morale than men. The estimate for the gender dummy is highly significant and lies at around -0.04, meaning that tax morale is about 4 percentage points lower for men relative to women. This effect remains significant and around the same magnitude as we include net income, risk and patience and political preferences as covariates. The literature on the gender-wage gap finds that gender differences might partly be driven by risk aversion (Bertrand 2011) and it might be the case that previously found gender effects in tax morale are also driven by omitted risk aversion. However, this does not seem to be the case: the gender difference in tax morale does not diminish once we condition on risk.<sup>29</sup> It

---

<sup>26</sup>For reasons of brevity, Table 3 presents the regression coefficients only, and not the (robust) standard errors. Significance stars indicate precision in this table. Table 7 in the Appendix presents the same regression table, but including (robust) standard errors.

<sup>27</sup>These variables have a slightly lower number of non-missing observations, which is why we add them subsequently.

<sup>28</sup>Doerrenberg and Peichl (2013) briefly summarize the survey literature with respect to drivers of tax morale.

<sup>29</sup>See the variable description in section 2 for a discussion on the reliability of our measure of risk

is thus an insight adding to previous literature that the previously found gender difference is not driven by risk aversion.<sup>30</sup>

A further strong driver of tax morale in our data is age; tax morale strongly increases with age. For example, tax morale of individuals older than 59, as well as of individuals between 50 and 59 years, is about 10-11 percentage points higher than for individuals younger than 30. The respective effect for age groups 30-39 and 40-49, relative to being younger than 30, is around 7 percentage points. These effects are all statistically significant and they are not driven by variables that are correlated with age and also potentially matter for tax morale; for example, neither retirement status, marital status, education (specification I) nor income (II) considerably weaken the effect. Patience, risk aversion (III) and political preferences (IV) do not diminish the age effect either. This finding is in line with the survey literature and hence seems to be very robust.<sup>31</sup> We further find an effect of retirement on tax morale, yet only significant in specification (IV). Being retired increases tax morale by about 4 percentage points – even conditional on age. We also see that household size is negatively correlated with tax morale. The effects of marital status and employment status on tax morale are not statistically significant. The effect of education loses significance as soon as income is added to the regression (specifications I and II). The difference in tax morale between East and West Germany is small and not statistically significant.

The effect of income (specification II) appears to be positive but is only statistically significant for the medium-high income group (those with household net income between 3,000 and 5,000 Euro). Being in this income group increases tax morale by about 4-6 percentage points, relative to the poorest households (significant across all specifications). The lack of significance for the other income groups might reflect the ambiguous theoretical effect of income on tax morale: Evasion yields higher returns for richer people, but they also have higher societal stakes and are more affected by sanctions (i.e., losing a well paid job). The empirical picture from surveys is also ambiguous. Some studies find insignificant effects (e.g., Konrad and Qari 2012 for Europe), while other studies find negative effects (e.g., Alm and Torgler 2006 for US and Europe).

Risk aversion and tax morale are strongly positively correlated; participants with high risk aversion are more likely to report higher tax morale. The magnitude of this effect is quite sizable: participants in the 4th and 5th category of the risk-aversion variable (with 5 indicating the highest risk aversion) report a tax morale that is 10-11 percentage

---

attitudes. With respect to average gender differences in risk, we indeed see in our data that women are more risk averse; average risk aversion for women is at 3.9 and at 3.5 for men.

<sup>30</sup>For illustration purposes, we show the unconditional means for men and women in Appendix Figure 5.

<sup>31</sup>The unconditional means for the different age groups are depicted in Figure 6 in the Appendix for illustration purposes.

points higher than for risk-loving participants (category 1).<sup>32</sup> While this relationship is intuitive, it has – to the best of our knowledge – not been clearly established in previous literature, probably due to reasons of data availability (tax morale and risk preferences are not measured together).<sup>33</sup>

The effect of patience is positive in specification (III) but vanishes as political preferences are included (IV). With regard to these political preferences, our regressions reveal that tax morale is significantly lower among right-wing participants. The estimates suggest that tax morale is about 10 percentage points lower for right-wingers than for conservatives.

The results in specification (V) show that none of the above results are confounded by the treatment information (i.e., adding a variable for the treatment group does not change any of the described regression results).

## 4 Results of randomized survey experiment

**Main results.** The main results of the experimental variation on tax morale are presented in Figure 1 and Table 4. Figure 1 shows the average levels of tax morale in each experimental group along with 95% confidence bands. Average tax morale is around 89% in the control group, 87% in the social-norm group and 90% in the reciprocity group. The p-values from pair-wise non-parametric Wilcoxon rank-sum (Mann-Whitney) tests are as follows: control vs social norm: 0.165; control vs reciprocity: 0.256; social norm vs reciprocity: 0.012. The social-norm treatment thus slightly decreased tax morale relative to the control group. Adding the reciprocity component to the social-norm information then significantly increases tax morale. Note, again, that our augmented treatment structure implies that we compare the reciprocity treatment relative to the social-norms treatment to isolate the effect of the reciprocity treatment. Comparing the reciprocity treatment to the control treatment does not identify the reciprocity effect because this comparison includes the effect of the social-norms treatment *and* the reciprocity treatment. The significant difference between the social-norms and reciprocity groups thus indicates a significant effect of the reciprocity treatment.

An (ex-post) power analysis, building on the standard deviations that we see in the data (with power  $(1 - \beta) = 0.8$  and type-I error rate of 0.1) reveals that we would have required a somewhat larger sample size in order to detect with statistical precision

---

<sup>32</sup>Figure 7 in the Appendix shows unconditional levels of tax morale by risk attitudes.

<sup>33</sup>The WVS in some waves includes a question about 'which things are most important if you were looking for a job'. Answer category 'A safe job with no risk of closing down or unemployment' is sometimes used to construct a measure of risk which is then included as an explanatory variable in tax-morale regressions (e.g., Torgler 2006). However, in light of the question's focus on job search and considering that even risk-averse people might prefer a safe job (many safe jobs are also very well paid, for example civil servants), we are uncertain if this question really captures risk aversion.

the difference between the *control* and *social norms* groups that we actually see in our data: we would have required 2745 observations per group in order to obtain statistically significant evidence (1-sided test) that the observed difference between the groups is not driven by chance. In line with the low p-value that we obtain for the difference between groups *social norms* and *reciprocity*, the power analysis yields that we would have required 833 observations per group (which is less than what we actually have) in order to get statistically significant evidence for the observed difference in tax morale between individuals in group *reciprocity* and group *social norms* (power  $(1 - \beta) = 0.8$ , one-sided test).<sup>34</sup>

The non-parametric findings are mirrored in the OLS regressions shown in Table 4. Specification (I) is a simple regression of tax morale on the treatment indicators, without the inclusion of any covariates. Not surprisingly, this regression simply reflects the non-parametric differences in means. The p-value from a t-test that compares the social-norm and reciprocity groups stands at 0.012 and is hence statistically significant. This confirms that the reciprocity treatment has a statistically significant effect on tax morale.

Specification (II) adds a basic set of pre-defined demographics as control variables.<sup>35</sup> The addition of control variables increases the coefficient measuring the difference between the control and social norm group by a little bit (to 2.2% percentage points) and makes the estimate slightly more precise; it is significant at the 10% level in this specification. That is, adding control variables, we now find a weak indication that the social norm treatment has a negative effect on tax morale. The difference between the social-norm and reciprocity groups is significant on the 1% level (p-value of 0.007) in this specification. This confirms the findings in specification (I) and the non-parametric tests that the reciprocity treatment has an effect on tax morale. Because of a few missing observations in the set of control variables, the number of observations in specification (II) is slightly smaller than in specification (I). Specification (III) estimates the effect of tax morale on the treatment indicators without any control variables, but based on the estimation

---

<sup>34</sup>Note that performing ex-ante power analyses during the design stage of our survey experiment was very difficult. Such analyses would have required reliable predictions of effects sizes, which is challenging in light of the fact that similar survey studies are very rare. In addition, we faced a given number of participants in the GIP and it would have been difficult to adjust the sample size in response to the results of an ex-ante power analysis. We therefore do not present the results of any ex-ante power analyses. However, based on an analysis of treatment-effect sizes in survey experiments, the review paper by Haaland et al. (2020) suggests that randomized survey experiments should have about 700 observations per treatment arm. The number of observations in our experimental treatment arms are in accordance with this suggestion.

<sup>35</sup>The demographics that are added as control variables in specification (II) are: gender, age, marital status, household size, employment status, retirement status, education, and a dummy indicating East Germany. Including the employment status as a control variable is particularly important because this variable is not fully balanced across experimental groups, as we show in section 2 (recall that the reason for this imbalance is the very low overall level of unemployment).

sample of specification (II) (which includes control variables). The treatment estimates suggest that increased precision in specification (II) is due to the addition of control variables and not caused by a different composition of the estimation sample.

**Discussion of main results.** We find two main results: (i) information about the general extent of tax evasion have a negative effect on tax morale, relative to the control group (though not statistically significant in the basic regression specification without control variables). (ii) If an appeal to reciprocity is added to the social-norm information, tax morale becomes significantly larger (significant effect of reciprocity relative to social norm) and even larger than in the control group.

How can these results be rationalized? The effect of the social-norm treatment, relative to control, suggests that manipulating the social norm of tax compliance through the provision of information about the commonness of evasion affects tax morale. This effect is in line with different strands of literature which find similar effects in different contexts; for example late tax payments, public good provision, charitable giving and energy saving (see the Introduction for references and details). Individuals are most likely very uncertain about the true extent of tax evasion. The negative effect of the tax-gap information suggests that participants perceived the tax-gap numbers presented in the information treatment to be considerably high (perceived initial tax evasion was unfortunately not surveyed).<sup>36</sup> So in line with for example Cialdini (2003), we confirm that a social-norm manipulation can backfire when it reveals a certain behavior as regrettably frequent. Simply speaking, the underlying mechanisms is something like "if so many others do it, it must be ok".<sup>37</sup>

The positive effect of adding the reciprocity component to the social-norm information (i.e., group reciprocity vs group social norm) indicates that an appeal to reciprocity affects tax morale (recall that, due to the augmented treatment structure, the difference between groups social norm and reciprocity isolates the effect of reciprocity). In addition, the reciprocity finding suggests that a potential social-norm backfire effect can be offset when the social-norm information are presented in a certain context and when the consequences of (not following) the social-norm are made salient. Relating the information about the tax gap to information about foregone tax-gap-induced government expenses makes participants realize that beneficial government services can only be provided *in return* for compliant tax payers. In the words of Feld and Frey (2007), the reciprocity component reminds participants of the 'implicit contract' between the government and

---

<sup>36</sup>Press coverage and anecdotal evidence tend to give the impression that tax evasion in Germany is not as much of a concern as in other countries (see the whole debate about tax evasion in Greece in the context of the Euro crisis).

<sup>37</sup>Or as phrased by Cialdini (2003): "Within the statement 'Many people are doing this undesirable thing' lurks the powerful and undercutting normative message 'Many people *are* doing this'."

the citizens and that this contract is threatened through tax evasion. Our finding also speaks to the recent paper by Besley (2020) which models a reciprocal relationship between citizens and the governments where the government provides public goods and citizens voluntarily pay taxes. The positive effect of the reciprocity treatment, relative to social-norm treatment, is in line with studies in the literature finding that reciprocity matters for behavior and that people are willing to give if they receive something in return (see Introduction for references and examples).

The size of the treatment effects is around 2-3 percentage points. In light of an average tax-morale level of 89%, this effect does not appear to be enormous. However, tax morale is usually seen to be a 'deep' parameter which is shaped over a lifetime by experiences as a taxpayer, perceptions of and attitudes towards the government, culture and social interaction with peers. This implies that it is likely to be fairly inelastic and small interventions can hardly have large effects. Our experimental manipulation consisted of only one or two additional sentences and was therefore fairly minor. In light of these considerations, the experiment-induced changes in tax morale in our study might be more important than it appears on first glance.<sup>38</sup>

Another way to assess the importance of the treatment effects is to consider the inverse of tax morale as a benchmark;  $(1 - TaxMorale)$  can be labeled as 'acceptance of tax evasion'. Our data show that only 11% of participants find tax evasion acceptable. Using this as the benchmark for assessing the magnitude of treatment effects in the range of 2-3 percentage points sheds a different light on the importance of the results and lets them appear quite sizable.

**Heterogeneity of treatment effects.** In a next step, we investigate if the experimental interventions had differential effects on different type of participants. For this purpose, we run OLS regressions of the following form separately for each covariate:  $TM_i = \beta_1 Treat_i + \beta_2 Covariate_i + \beta_3 (Treat_i \times Covariate_i) + \epsilon_i$ . The outcome variable  $TM_i$  is tax morale of participant  $i$ ,  $Treat_i$  indicates treatment dummies,  $Covariate_i$  is a covariate, and  $Treat_i \times Covariate_i$  is a full interaction between the treatment dummies and the categories of the respective covariate.

The results are presented in table 5. For reasons of brevity, we do not report the heterogeneous effect of each covariate (available upon request), but only for those covariates where we find some significant heterogeneity. Again for brevity, the table only reports the regression coefficients of the interaction terms,  $Treat_i \times Covariate_i$ . The coefficients for  $Treat_i$  and  $Covariate_i$ , as well as standard errors, are not reported

---

<sup>38</sup>On a related note, Luttmer and Singhal (2014) stress that small or even null findings of some field experiments might be due to the weak strength of the experimental manipulation and the "deep" attitudes that are behind compliance behavior. They argue that this should not necessarily be interpreted as evidence that a certain mechanism cannot be powerful.

in the table (significance stars based on robust standard errors and the usual levels of significance).

Overall, we do not find much heterogeneity of the treatments effects.<sup>39</sup> As the table shows, younger age groups respond stronger to the interventions than older age groups. This might be due to lower average tax morale in the younger groups, implying that there is more room for an increase in tax morale. We further find that married participants respond somewhat stronger to the social-norm treatment than unmarried ones. We also find that participants living in large households respond stronger to the social-norm treatment. These results might indicate that social norms have differential effects depending on the social ties and environment of the participants. Household size also has an effect on the response to the reciprocity treatment; the effect of reciprocity is more negative for larger households.

Interestingly, we find that East German respondents respond significantly less to the two treatment interventions than West German respondents. This source of heterogeneity might be related to the well documented finding that the political culture, preferences and social norms in East Germany are still different than in West Germany (e.g., Alesina and Fuchs-Schuendeln 2007; Brosig-Koch et al. 2011; Fuchs-Schuendeln and Masella 2016.) As a result, East Germans might respond differently to information about other people's behavior and matters of government intervention.

We further find one heterogeneous effect of income; participants in the third income group (net household income between 3,000 and 5,000 Euro) respond more strongly to the social-norm treatment. If we use a binary version of the income-group variable (with '0' for household incomes less than 3,000 Euro and '1' for more than 3,000 Euro), we see that the richer households respond stronger to both the social-norm and reciprocity treatment (p-values of interaction terms: 0.003 and 0.060. These results for heterogeneity w.r.t. income are not reported in the regression table for reasons of brevity.)

In their randomized survey experiment on redistributive preferences, Alesina et al. (2018) find heterogeneous effects with respect to political preferences. Accordingly, we would maybe expect that, for example, left-leaning respondents respond more to the reciprocity treatment in our experiment because they have a higher preference for government spending. We indeed find that left-wing participants respond slightly more positive to the reciprocity treatment (interaction coefficient of 0.076). This coefficient is just beyond conventional significance though (p-value: 0.120). All other interactions of the treatment indicators with the political categories are far from conventional significance levels. Heterogeneity with respect to other (not reported) covariates are not significant either.

---

<sup>39</sup>We can, of course, not rule out that this analysis of heterogeneous responses lacks statistical power as the sample sizes for the different groups can get small.

**Robustness of main results.** We used OLS regressions in all previous analyses. Appendix Table 8 presents the results from probit regressions with tax morale as the dependent variable (the table is equivalent to the main OLS regression table 4, but using probit regressions). The results are fully in line with the previous OLS regressions. We observe negative coefficients for the social-norm treatment and positive coefficients for the reciprocity group, both relative to the control group. As in the OLS regressions, the effect of social norms is statistically significant in specifications (II) where covariates are added to the regression specification.

Importantly, the difference between the social-norm estimate and the reciprocity estimate is statistically significant in all specifications – as indicated by the p-values for this difference which are reported in the table. That is, we observe significantly higher tax morale in the reciprocity group relative to the social-norm group, and thus a positive effect of reciprocity.

We use a simple tax morale dummy as the outcome in all preceding analyses. Our second robustness test considers differences across experimental groups in the 6-pt scale version of the tax-morale question. Simple  $chi^2$  tests indicate that there neither are significant differences in tax morale (measured on 6pt scale) between control group and social-norm group (p-value: 0.502) nor between control group and reciprocity group (p-value: 0.651). However, we do find a significant difference in the 6-pt version of tax morale when we compare the social-norms group to the reciprocity group (p-value: 0.027). This latter result confirms our previous findings (using the tax-morale dummy) that the reciprocity treatment has a significant effect on tax morale (recall, again, the augmented treatment structure which implies that the reciprocity effect is identified by comparing the reciprocity group relative to the social-norm group).

As discussed in Section 2, the binary dummy version of the tax-morale variable accounts for the distribution of the original 6-point variable. The first two categories of the original variable (*very justifiable* and *justifiable*) were only selected by 4.65% of all survey participants. This distribution likely explains why statistical precision is somewhat smaller when we use the 6-point scale version of the variable.<sup>40</sup>

## 5 Concluding remarks

We study intrinsic motivations for tax compliance in the context of a randomized survey experiment. We integrate a commonly used question on tax morale into a representative survey in Germany and combine it with randomized information treatments. The first contribution of our paper is to shed new light on the anatomy of intrinsic motivations. We

---

<sup>40</sup>As we also discuss in Section 2, using a recoded version of the original tax-morale question is for the same reason the usual procedure in the literature using the World Values Survey.

confirm earlier findings on the (correlational) effects of gender and age on tax morale. We further show that these previous findings are not confounded by risk aversion or patience, and find that risk aversion and tax morale are positively correlated. Participants with right-wing political attitudes have lower tax morale.

In light of mixed findings in the literature on the role of social norms and reciprocity for tax compliance, our main contribution is to provide new evidence on this role. To do so, we conduct one of the first randomized survey experiments in the context of tax compliance. Our customized survey experiment allows us to study social norms and reciprocity within the same experimental design. In our experimental interventions, we (i) inform people about the extent of tax evasion in industrialized countries and (ii) make it salient that the tax-evasion-induced foregone revenue has high consequences for the provision of public goods through the government. That is, treatment (i) manipulates the social norm of tax compliance and treatment (ii) adds a reciprocity component by reminding participants that tax compliance and government services are closely linked. We particularly find that the appeal to reciprocity increases tax morale. In light of the usual perception that tax-morale attitudes are fairly inelastic and considering the 'acceptance of tax evasion' as a benchmark, the size of the treatment effects appears fairly sizable. Our results have important policy implications. Adding elements of reciprocity and highlighting the use of tax revenues might offer easy opportunities for fighting tax evasion.

## References

- Alesina, A. and N. Fuchs-Schuendeln (2007). Goodbye lenin (or not?): The effect of communism on people’s preferences. *American Economic Review* 97(4), 1507–1528.
- Alesina, A., S. Stantcheva, and E. Teso (2018). Intergenerational mobility and support for redistribution. *American Economic Review* 108(2), 521 – 554.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics* 95(9), 1082 – 1095.
- Allingham, M. G. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1(3-4), 323 – 338.
- Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance* 19(1), 54–77.
- Alm, J. and B. R. Jackson (1993). Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior and Organization* 22, 285 – 303.
- Alm, J. and A. Malezieux (2020). 40 years of tax evasion games: A meta-analysis. *Experimental Economics*. forthcoming.
- Alm, J. and B. Torgler (2006). Culture differences and tax morale in the United States and in Europe. *Journal of Economic Psychology* 27(2), 224 – 246.
- Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature* 36(2), 818–860.
- Andreoni, J. and J.-K. Scholz (1998). An econometric analysis of charitable giving with interdependent preferences. *Economic Inquiry* 36(3), 410–428.
- Bergolo, M. L., R. Ceni, G. Cruces, M. Giacobasso, and R. Perez-Truglia (2017). Tax audits as scarecrows: Evidence from a large-scale field experiment. NBER Working Paper 23631.
- Bertrand, M. (2011). New perspectives on gender. *Handbook of Labor Economics* 4(Part B), 1543 – 1590.
- Besley, T. (2020). State capacity, reciprocity, and the social contract. *Econometrica* 88(4), 1307–1335.
- Besley, T. J., A. Jensen, and T. Persson (2015). Norms, Enforcement, and Tax Evasion. CEPR Discussion Papers 10372.
- Blom, A. G., D. Bossert, F. Gebhard, J. Herzing, and U. Krieger (2016). SFB 884 ‘Political Economy of Reforms’: German Internet Panel, wave 14 (November 2014). GESIS data archive, Cologne. ZA5925, Data file version 2.0.0.

- Blom, A. G., C. Gathmann, and U. Krieger (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods* 27(4), 391–408.
- 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(1), 125 – 138.
- Bohne, A. and J. S. Nimczik (2018). Information frictions and learning dynamics: Evidence from tax avoidance in ecuador. IZA discussion paper no. 11536.
- Bott, K. M., A. W. Cappelen, E. O. Sorensen, and B. Tungodden (2019). You’ve got mail: A randomized field experiment on tax evasion. *Management Science*. forthcoming.
- Brosig-Koch, J., C. Helbach, A. Ockenfels, and J. Weimann (2011). Still different after all these years: Solidarity behavior in east and west germany. *Journal of Public Economics* 95(11), 1373 – 1376.
- Castro, L. and C. Scartascini (2015). Tax compliance and enforcement in the pamapas evidence from a field experiment. *Journal of Economic Behavior & Organization* 116, 65 – 82.
- Cebula, R. J. (2013). New and current evidence on determinants of aggregate federal personal income tax evasion in the united states. *American Journal of Economics and Sociology* 72(3), 701–731.
- Chetty, R., J. N. Friedman, and E. Saez (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review* 103(7), 2683–2721.
- Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science* 12(4), 105–109.
- Congdon, W. J., J. R. Kling, and S. Mullainathan (2009). Behavioral Economics and Tax Policy. *National Tax Journal* 62(3), 375–386.
- Corneo, G. and H. P. Gruener (2002). Individual preferences for political redistribution. *Journal of Public Economics* 83(1), 83 – 107.
- Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics* 98, 100 – 112.
- Cullen, J. B., N. Turner, and E. L. Washington (2018). Political alignment, attitudes toward government and tax evasion. NBER Working Paper 24323.

- Cummings, R. G., J. Martinez-Vazquez, M. McKee, and B. Torgler (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior & Organization* 70(3), 447 – 457.
- de Quidt, J., J. Haushofer, and C. Roth (2018, November). Measuring and bounding experimenter demand. *American Economic Review* 108(11), 3266–3302.
- DeBacker, J., B. T. Heim, and A. Tran (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the united states. *Journal of Financial Economics* 117(1), 122 – 138.
- Doerrenberg, P. and A. Peichl (2013). Progressive taxation and tax morale. *Public Choice* 155(3-4), 293–316.
- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: measurement, determinants and behavioral consequences. *Journal of the European Economic Association* 9(3), 522–550.
- Dolls, M. and N. Wehrhoefer (2018). Attitudes towards euro area reforms: Evidence from a randomized survey experiment. CESifo working paper series no. 7141.
- Dulleck, U., J. Fooker, C. Newton, A. Ristl, M. Schaffner, and B. Torgler (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics* 134, 9–18.
- Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy* 8(3), 203–32.
- Engelmann, D., E. Janeba, L. Mechtenberg, and N. Wehrhoefer (2018). Preferences over taxation of high income individuals: Evidence from survey and laboratory experiments. mimeo.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*. Forthcoming.
- Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA Discussion Paper 9674.
- Falkinger, J. (1988). Tax Evasion and Equity: A Theoretical Analysis. *Public Finance = Finances publiques* 43(3), 388–395.
- Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14(3), 159–181.
- Feld, L. P. and B. S. Frey (2007). Tax compliance as the result of a psychological tax contract: The role of incentives and responsive regulation. *Law & Policy* 29(1), 102 – 120.

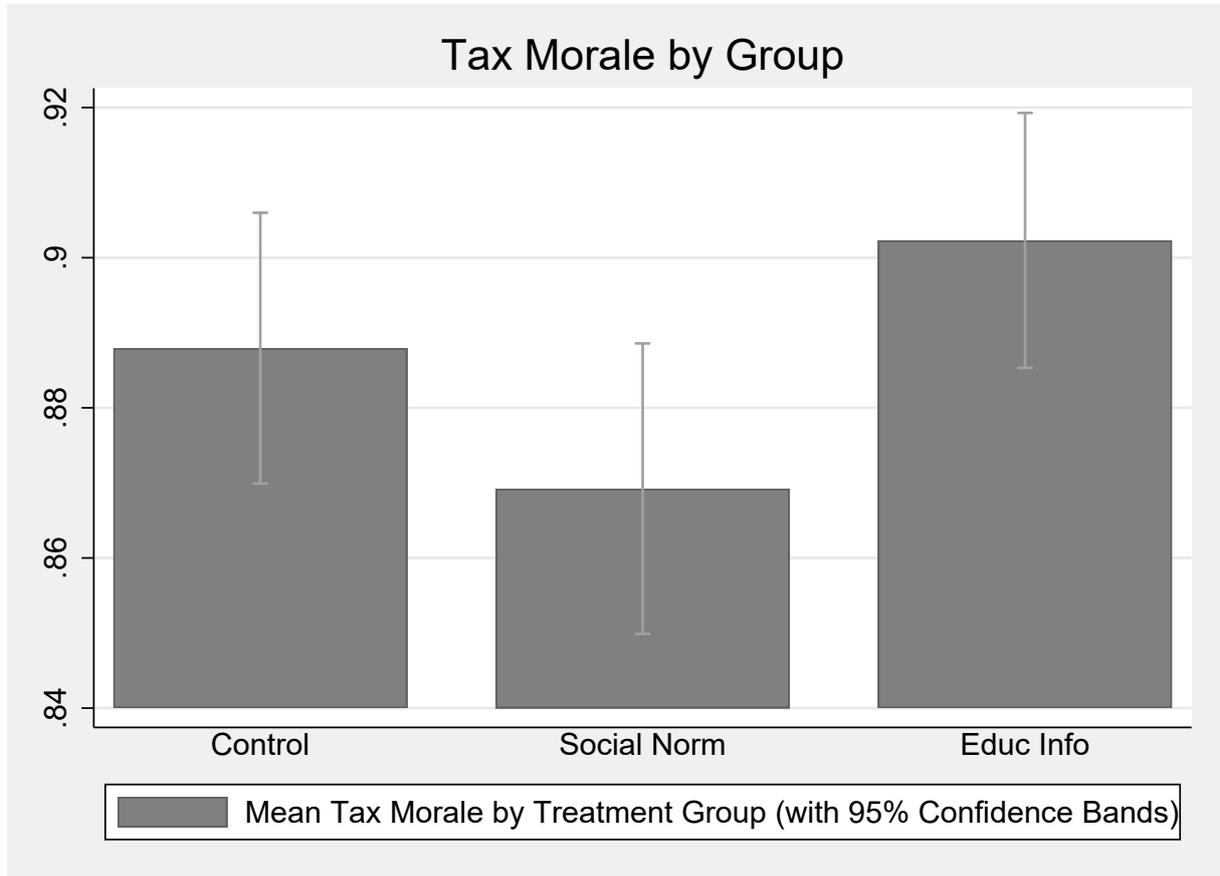
- Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing Enforcement Strategies In The Field: Threat, Moral Appeal And Social Information. *Journal of the European Economic Association* 11(3), 634–660.
- Fong, C. M. and E. F. Luttmer (2011). Do fairness and race matter in generosity? evidence from a nationally representative charity experiment. *Journal of Public Economics* 95(5-6), 372 – 394.
- Fortin, B., G. Lacroix, and M.-C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics* 91(11), 2089 – 2112.
- Frey, B. S. and S. Meier (2004). Social comparisons and pro-social behavior: Testing ”conditional cooperation” in a field experiment. *American Economic Review* 94(5).
- Frey, B. S. and B. Torgler (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics* 35(1), 136 – 159.
- Frimmel, W., M. Halla, and J. Paetzold (2018). The Intergenerational Causal Effect of Tax Evasion: Evidence from the Commuter Tax Allowance in Austria. *Journal of the European Economic Association* 17(6), 1843–1880.
- Fuchs-Schuendeln, N. and P. Masella (2016). Long-lasting effects of socialist education. *The Review of Economics and Statistics* 98(3), 428–441.
- Haaland, I., C. Roth, and J. Wohlfart (2020). Designing information provision experiments. CAGE working paper no. 484.
- Halla, M. (2012). Tax morale and compliance behavior: First evidence on a causal link. *The B.E. Journal of Economic Analysis & Policy* 12(1).
- Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behaviorist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics* 148, 14 – 31.
- IRS (2007). Reducing the federal tax gap. a report on improving voluntary compliance. Report by US Internal Revenue Service (IRS), online at: [https://www.irs.gov/pub/irs-news/tax\\_gap\\_report\\_final\\_080207\\_linked.pdf](https://www.irs.gov/pub/irs-news/tax_gap_report_final_080207_linked.pdf).
- Kerschbamer, R. and D. Müller (2017). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. University of insbruck working paper 2017-16.
- Keser, C. and F. Van Winden (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics* 102(1), 23–39.
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica* 79(3), 651 – 692.

- Konrad, K. A. and S. Qari (2012). The last refuge of a scoundrel? Patriotism and tax compliance. *Economica* 79(315), 516–533.
- Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review* 105(4), 1478–1508.
- Lamberton, C., J.-E. D. Neve, and M. I. Norton (2014). Eliciting taxpayer preferences increases tax compliance. Harvard Business School Working Paper 14-106.
- List, J. A. and D. Lucking-Reiley (2002). The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. *Journal of Political Economy* 110(1), 215–233.
- Luttmer, E. F. P. (2001). Group loyalty and the taste for redistribution. *Journal of Political Economy* 109(3), 500–528.
- Luttmer, E. F. P. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives* 28(4), 149–68.
- Müller, D. and S. Renes (2017). Fairness views and political preferences - evidence from a large online experiment. University of insbruck working paper 2017-10.
- Mummolo, J. and E. Peterson (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*. forthcoming.
- Ortega, D., L. Ronconi, and P. Sanguinetti (2016). Reciprocity and willingness to pay taxes:: Evidence from a survey experiment in latin america. *Economia* 16(2), 55–87.
- Paetzold, J. and H. Winner (2016). Taking the high road? compliance with commuter tax allowances and the role of evasion spillovers. *Journal of Public Economics* 143, 1 – 14.
- Perez-Truglia, R. and U. Troiano (2015). Shaming tax delinquents. NBER Working Paper 21264.
- Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review* 105(8), 2539–69.
- Richardson, G. (2006). Determinants of tax evasion: A cross-country investigation. *Journal of International Accounting, Auditing and Taxation* 15(2), 150 – 169.
- Scholz, J. T. and M. Lubell (1998). Trust and taxpaying: Testing the heuristic approach to collective action. *American Journal of Political Science* 42(2), 398–417.
- Sjoberg, F. M., J. Mellon, T. Peixoto, J. Hemker, and L. L. Tsai (2019). Voice and punishment. a global survey experiment on tax morale. World Bank policy research working paper 8855.

- Slemrod, J. (2003). Trust in public finance. In S. Cnossen and H.-W. Sinn (Eds.), *Public Finance and Public Policy in the New Century*, pp. 49 – 88. Cambridge, USA: The MIT Press.
- Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives* 21(1), 25–48.
- Torgler, B. (2006). The importance of faith: Tax morale and religiosity. *Journal of Economic Behavior & Organization* 61(1), 81 – 109.
- Torgler, B. and F. Schneider (2009). The impact of tax morale and institutional quality on the shadow economy. *Journal of Economic Psychology* 30(2), 228 – 245.
- Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy* 26(1), 89 – 103.
- Vischer, T., T. Dohmen, A. Falk, D. Huffman, J. Schupp, U. Sunde, and G. G. Wagner (2013). Validating an ultra-short survey measure of patience. *Economics Letters* 120(2), 142 – 145.
- Weimann, J. (1994). Individual behaviour in a free riding experiment. *Journal of Public Economics* 54(2), 185 – 200.
- Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics* 6(3), 299–310.
- Zukunft durch Bildung (2011). Zukunft durch bildung – deutschland will’s wissen. Survey on education by roland berger strategy consultants, bertelsmann stiftung, bild and hurriyet, summary online at: <http://www.bildung2011.de/download/Ergebnisse-der-Online-Buergerbefragung.pdf>.

# Tables and Figures

Figure 1: Tax Morale by Experimental Group



Notes: Average Tax Morale by experimental group with 95% confidence bars. The outcome variable is survey-based tax morale as described in Section 2. Treatment groups as described in Section 2. The treatment structure is augmented, implying that the reciprocity group includes the reciprocity treatment and the social-norms treatment. Total number of observations is 3525 with even distribution across experimental groups. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 1: Overview of variables

Variable	Measurement	Orig. question
Tax Morale	(1): Evasion is 'not at all justifiable', 'not justifiable' or 'rather not justifiable'; (0): Evasion is 'very justifiable', 'justifiable' or 'rather justifiable'	<i>CF14015</i>
Treatment status	(1): Control; (2) Social norm; (3) Reciprocity	<i>expCF1401</i>
Gender	(1): Male; (0): Female	<i>gender 14</i>
Age	(1): < 30; (2): 30-39; (3): 40-49; (4): 50-59; (5): > 59	<i>age cat 14</i>
Marital Status	(1): Married; (0): Not married	<i>marital status 14</i>
Household size	(1): 1; (2) 2; (3) 3; (4): 4; (5): > 4	<i>number hh members 14</i>
Employment status	(1): Unemployed; (0): Employed	<i>occupation 14</i>
Retirement Status	(1): Retired; (0): Not retired	<i>occupation 14</i>
East Germ.	(1): Lives in East Germany; (0): West Germany	<i>state</i>
Household Income (net)	(1): 0-1500 Euro; (2) 1500-3000 Euro; (3) 3000-5000 Euro; (4): > 5000	<i>AA1305x</i>
Risk Aversion	Own risk perception measured on 11-point scale. We recode the variable to have 5 categories from (1) risk loving to (5) risk averse	<i>ZE14074</i>
Patience	Own perception of patience measured on 11-point scale. We recode the variable to have 5 categories from (1) not patient to (5) patient	<i>AE14007</i>
Political preference	(1): Conservative; (2): Moderate left; (3): Right wing; (4) Left wing	<i>CE14140</i>
Education	(1): no degree; (2) high school without university qualification; (3) high school with university qualification or apprenticeship combined with high school without university qualification; (4): apprenticeship and high school degree with university qualification; (5): University degree or more	<i>educ school 14</i>

Notes: Overview of all variables used throughout the paper. We list the question number in the original GIP survey in the last column (*Orig. question*). All variables come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 2: Summary statistics

Variable	N	mean	sd	min	max	p50
Tax morale	3525	0.89	0.32	0.00	1.00	1.00
Control	3532	0.33	0.47	0.00	1.00	0.00
Social norm	3532	0.33	0.47	0.00	1.00	0.00
Reciprocity	3532	0.33	0.47	0.00	1.00	0.00
Gender	3574	0.49	0.50	0.00	1.00	0.00
Married	3575	0.58	0.49	0.00	1.00	1.00
Retired	3575	0.16	0.37	0.00	1.00	0.00
Unemployed	3575	0.03	0.16	0.00	1.00	0.00
Household size	3571	2.54	1.13	1.00	5.00	2.00
East Germ.	3575	0.20	0.40	0.00	1.00	0.00
Age < 30	3573	0.18	0.38	0.00	1.00	0.00
Age 30-39	3573	0.16	0.37	0.00	1.00	0.00
Age 40-49	3573	0.19	0.40	0.00	1.00	0.00
Age 50-59	3573	0.23	0.42	0.00	1.00	0.00
Age > 60	3573	0.24	0.43	0.00	1.00	0.00
Conservative	2676	0.36	0.48	0.00	1.00	0.00
Moderate left	2676	0.44	0.50	0.00	1.00	0.00
Right wing	2676	0.11	0.31	0.00	1.00	0.00
Left wing	2676	0.09	0.29	0.00	1.00	0.00
Educ low	3574	0.01	0.09	0.00	1.00	0.00
Educ low-med	3574	0.08	0.27	0.00	1.00	0.00
Educ med	3574	0.52	0.50	0.00	1.00	1.00
Educ high-med	3574	0.15	0.35	0.00	1.00	0.00
Educ high	3574	0.24	0.43	0.00	1.00	0.00
Inc low	2919	0.15	0.36	0.00	1.00	0.00
Inc low-med	2919	0.40	0.49	0.00	1.00	0.00
Inc med	2919	0.35	0.48	0.00	1.00	0.00
Inc high	2919	0.10	0.30	0.00	1.00	0.00
Risk aversion	3517	3.67	1.12	1.00	5.00	4.00
Patience	3516	3.48	1.24	1.00	5.00	4.00

Notes: Summary Statistics for all variables. All variables are defined as described in Table 1. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 3: Anatomy of Tax Morale

Variable	Dep. var.: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
<u>Gender.</u> Reference category: <i>Female</i>					
Male	-0.041***	-0.042***	-0.038***	-0.035***	-0.035***
<u>Age.</u> Reference category: <i>&lt;30</i>					
30-39	0.063***	0.064**	0.061**	0.072**	0.070**
40-49	0.075***	0.089***	0.086***	0.076***	0.074**
50-59	0.113***	0.117***	0.116***	0.097***	0.096***
>59	0.116***	0.111***	0.108***	0.096***	0.095***
<u>Marital status.</u> Reference category: <i>Not Married</i>					
Married	-0.001	0.007	0.004	-0.002	-0.001
<u>Size of household.</u> Reference category: <i>1</i>					
2	-0.006	-0.019	-0.018	-0.017	-0.018
3	-0.029	-0.052**	-0.052**	-0.038	-0.040
4	-0.029	-0.047*	-0.047*	-0.051	-0.052
>4	-0.032	-0.051	-0.058*	-0.050	-0.052
<u>Unemployment.</u> Reference category: <i>Employed</i>					
Unemployed	-0.025	-0.031	-0.039	-0.045	-0.044
<u>Retirement Status.</u> Reference category: <i>Not Retired</i>					
Retired	0.012	0.030	0.029	0.035*	0.034*
<u>Education.</u> Reference category: <i>Low Education</i>					
2	0.057	0.065	0.059	-0.047	-0.045
3	0.099	0.085	0.078	-0.015	-0.012
4	0.131*	0.126	0.117	0.014	0.019
High Educ	0.145**	0.129	0.125	0.020	0.023
<u>East Germany.</u> Reference category: <i>West Germany</i>					
East Germany	-0.001	0.014	0.015	0.022	0.022
<u>Net household income.</u> Reference category: <i>Poor</i>					
2		0.017	0.014	0.037	0.037
3		0.045**	0.044**	0.062**	0.062**
rich		0.016	0.018	0.042	0.041
<u>Risk aversion.</u> Reference category: <i>Risk loving</i>					
2			0.058	0.060	0.060
3			0.089*	0.093*	0.093*
4			0.118**	0.109**	0.109**
risk averse			0.108**	0.101*	0.101*
<u>Patience.</u> Reference category: <i>Not patient</i>					
2			0.041	0.016	0.017
3			0.029	0.015	0.015
4			0.033	0.029	0.028
patient			0.070**	0.055	0.055
<u>Political preferences.</u> Reference category: <i>Conservative</i>					
social				0.007	0.007
right wing				-0.096***	-0.097***
left wing				-0.013	-0.012
<u>Experimental Treatment Group.</u> Reference category: <i>Control</i>					
Social Norm					-0.012
Reciprocity					0.014
constant	0.732***	0.715***	0.586***	0.700***	0.699***
N	3519	2881	2875	2236	2236
R2	0.036	0.039	0.049	0.054	0.056

Notes: The table presents the determinants of Tax Morale. OLS Regressions of Tax Morale on various covariates. Each column (I)-(V) presents the results of one regression with different sets of covariates. All variables are defined as described in Table 1. For reasons of brevity, (robust) standard errors are not reported. Significance stars indicate precision obtained from robust standard errors: \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. The same table including the presentation of robust standard errors is depicted in Appendix Table 7. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 4: Effect of experimental intervention on tax morale

Experimental Group.	Dep. Variable: Tax Morale		
	(I)	(II)	(III)
Reference category: <i>Control</i>			
Social Norm	-0.019 (0.013)	-0.022* (0.013)	-0.019 (0.014)
Reciprocity	0.014 (0.013)	0.013 (0.012)	0.014 (0.013)
constant	0.888*** (0.009)	0.737*** (0.078)	0.888*** (0.089)
p-val Norm vs Recipr.	0.012**	0.007***	0.012**
N	3525	3519	3519
R2	0.002	0.038	0.002
Demographics	no	yes	no

Notes: The table presents the effects of the randomized treatment interventions on Tax Morale. OLS Regressions of Tax Morale on treatment dummies. The experimental groups are: Control group, Social-norm group and Reciprocity group. Control is omitted, implying that the effects are relative to the Control Group. Line *p-val Norm vs Recipr.* presents the p-values from t-tests which compare if the regression coefficient for the Social-norm group is different from the regression coefficient for the Reciprocity group. The treatment structure is augmented. To isolate the effect of the reciprocity treatment, one thus needs to compare the reciprocity treatment relative to the social-norm treatment. Columns (I) and (II) differ in the included sets of covariates. (I): no covariates, (II): gender, age, marital status, household size, employment status, retirement status, education and a dummy indicating East Germany. Column (III) does not include covariates, but is estimated based on the estimation sample in column (II). All variables are defined as described in Table 1. Robust Standard Errors in Parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 5: Heterogeneous effects of experimental interventions

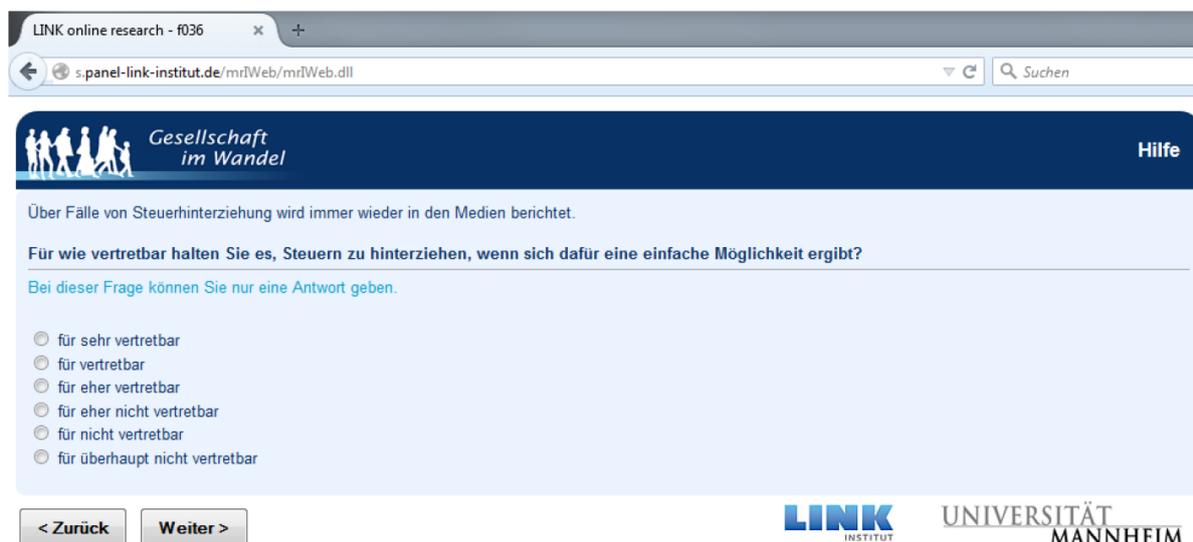
	Dep. var.: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
<u>Age.</u> Reference category: < 30					
Norm × 30-39	0.115**				
Norm × 40-49	0.097*				
Norm × 50-59	0.061				
Norm × > 59	0.031				
Recipr. × 30-39	0.085*				
Recipr. × 40-49	0.064				
Recipr. × 50-59	0.010				
Recipr. × > 59	0.020				
<u>Marital status.</u> Reference category: <i>Not married</i>					
Norm × Married		0.050*			
Recipr. × Married		0.015			
<u>Size of Household.</u> Reference category: 1					
Norm × HHsize 2			-0.008		
Norm × HHsize 3			-0.028		
Norm × HHsize 4			-0.030		
Norm × HHsize > 4			0.144**		
Recipr. × HHsize 2			-0.067**		
Recipr. × HHsize 3			-0.090**		
Recipr. × HHsize 4			-0.091**		
Recipr. × HHsize > 4			-0.047		
<u>Political preferences.</u> Reference category: <i>Conservative</i>					
Norm × social				-0.008	
Norm × right wing				0.029	
Norm × left wing				-0.020	
Recipr. × social				0.003	
Recipr. × right wing				0.029	
Recipr. × left wing				0.076	
<u>Eastern Germany.</u> Reference category: <i>West Ger.</i>					
Norm × East					-0.097***
Recipr. × East					-0.067**
N	3523	3525	3522	2654	3525
R2	0.028	0.006	0.012	0.014	0.003

Notes: Heterogeneous effects of the experimental interventions. Reported are coefficients of OLS regressions of the following form (which are estimated separately for each covariate):  $TM_i = \beta_1 Treat_i + \beta_2 Covariate_i + \beta_3 (Treat_i \times Covariate_i) + \epsilon_i$ . The outcome variable  $TM_i$  is tax morale of participant  $i$ ,  $Treat_i$  indicates treatment dummies,  $Covariate_i$  is a covariate, and  $Treat_i \times Covariate_i$  is a full interaction between the treatment dummies and the categories of the respective covariate. Specifications (I)-(V) present heterogeneous effects of different covariates. For reasons of brevity, estimates for heterogeneous effects of additional covariates are not displayed if no significant interactions found (available upon request). The treatment groups are: control, social-norm (*Norm*) and reciprocity (*Recipr.*). All variables are defined as described in Table 1. Robust standard errors not displayed for reasons of brevity. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

# Appendix

## Additional Figures and Tables

Figure 2: Screenshot of survey experiment: control group



Notes: The figure shows a screenshot of the survey question which is used to measure tax morale. This screenshot shows the screen which is shown to participants in the experimental *control group*. See section 2 for a description of the survey and the randomized survey experiment. German Internet Panel (GIP), wave 14. Sources: [http://reforms.uni-mannheim.de/internet\\_panel/Questionnaires/](http://reforms.uni-mannheim.de/internet_panel/Questionnaires/) and (Blom et al. 2016).

Figure 3: Screenshot of survey experiment: social-norm group



Notes: The figure shows a screenshot of the survey question which is used to measure tax morale. This screenshot shows the screen which is shown to participants in the experimental *social-norm group*. See section 2 for a description of the survey and the randomized survey experiment. German Internet Panel (GIP), wave 14. Sources: [http://reforms.uni-mannheim.de/internet\\_panel/Questionnaires/](http://reforms.uni-mannheim.de/internet_panel/Questionnaires/) and (Blom et al. 2016).

Figure 4: Screenshot of survey experiment: reciprocity group



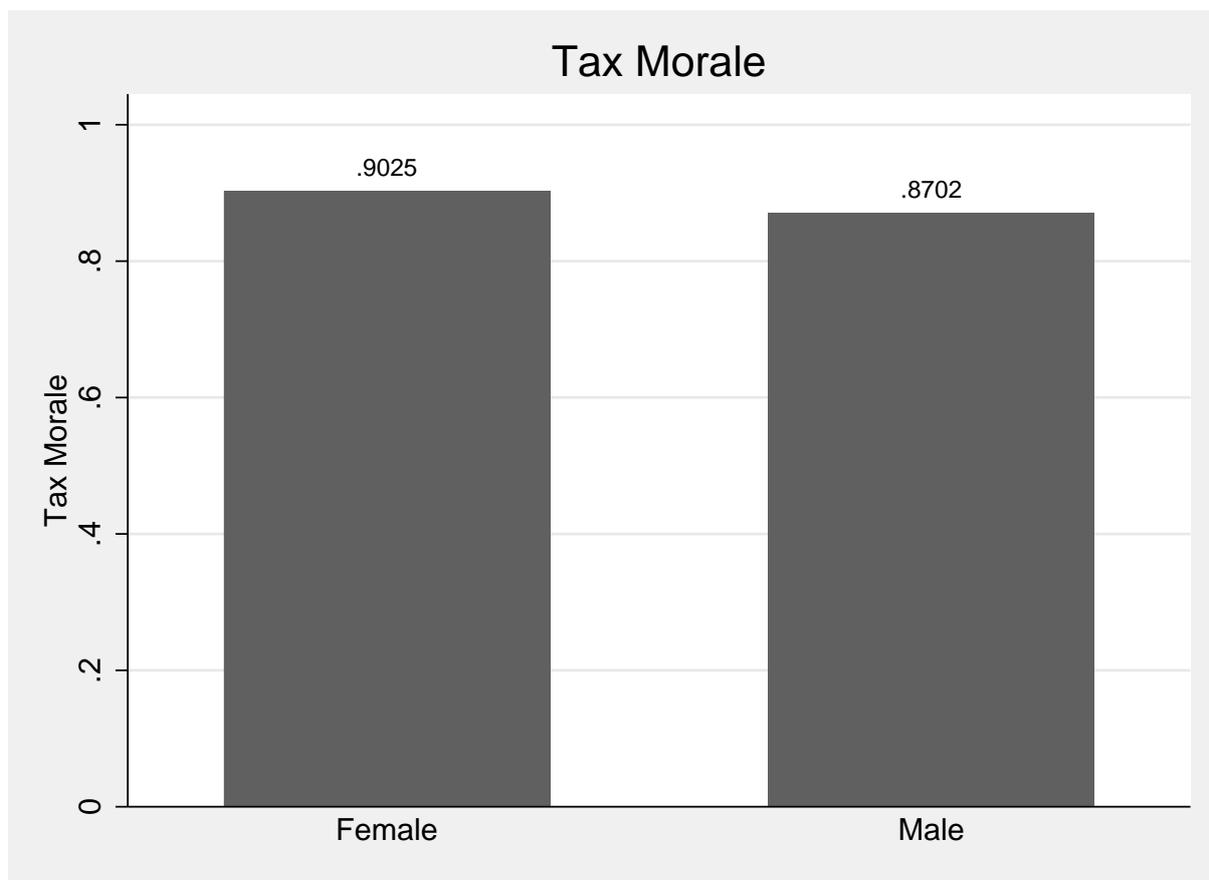
Notes: The figure shows a screenshot of the survey question which is used to measure tax morale. This screenshot shows the screen which is shown to participants in the experimental *reciprocity group*. See section 2 for a description of the survey and the randomized survey experiment. German Internet Panel (GIP), wave 14. Sources: [http://reforms.uni-mannheim.de/internet\\_panel/Questionnaires/](http://reforms.uni-mannheim.de/internet_panel/Questionnaires/) and (Blom et al. 2016).

Table 6: Randomization checks

	(I)	(II)	(III)
	Dep. var.: Treatment Group Indicator		
Variable	Control	Social Norm	Reciprocity
<u>Gender.</u> Reference category: <i>Female</i>			
Male	0.008 (0.016)	-0.013 (0.016)	0.005 (0.016)
<u>Age.</u> Reference category: <i>&lt;30</i>			
30-39	-0.027 (0.028)	-0.008 (0.027)	0.036 (0.027)
40-49	-0.041 (0.026)	-0.000 (0.026)	0.041 (0.026)
50-59	-0.017 (0.025)	-0.002 (0.025)	0.019 (0.025)
>59	-0.022 (0.025)	-0.001 (0.025)	0.022 (0.025)
<u>Marital status.</u> Reference category: <i>Not Married</i>			
Married	-0.023 (0.016)	0.015 (0.016)	0.009 (0.016)
<u>Size of household.</u> Reference category: <i>1</i>			
2	0.011 (0.023)	-0.027 (0.023)	0.016 (0.023)
3	-0.017 (0.027)	-0.022 (0.027)	0.039 (0.027)
4	-0.023 (0.028)	-0.002 (0.028)	0.025 (0.028)
>4	0.064* (0.038)	-0.057 (0.037)	-0.007 (0.037)
<u>Unemployment.</u> Reference category: <i>Employed</i>			
Unemployed	-0.130*** (0.043)	0.115** (0.052)	0.015 (0.050)
<u>Retirement Status.</u> Reference category: <i>Not Retired</i>			
Retired	-0.035* (0.021)	0.015 (0.022)	0.020 (0.022)
<u>Education.</u> Reference category: <i>Low Education</i>			
2	0.062 (0.084)	-0.054 (0.088)	-0.008 (0.090)
3	0.069 (0.080)	-0.022 (0.085)	-0.047 (0.086)
4	0.046 (0.082)	0.016 (0.087)	-0.063 (0.088)
high educ	0.020 (0.081)	0.010 (0.086)	-0.030 (0.087)
<u>Net household income.</u> Reference category: <i>Poor</i>			
2	-0.002 (0.026)	-0.010 (0.027)	0.012 (0.026)
3	0.015 (0.027)	-0.022 (0.027)	0.007 (0.027)
rich	-0.049 (0.034)	0.000 (0.036)	0.049 (0.036)
<u>East Germany.</u> Reference category: <i>West Germ.</i>			
	-0.050*** (0.019)	0.019 (0.020)	0.031 (0.020)

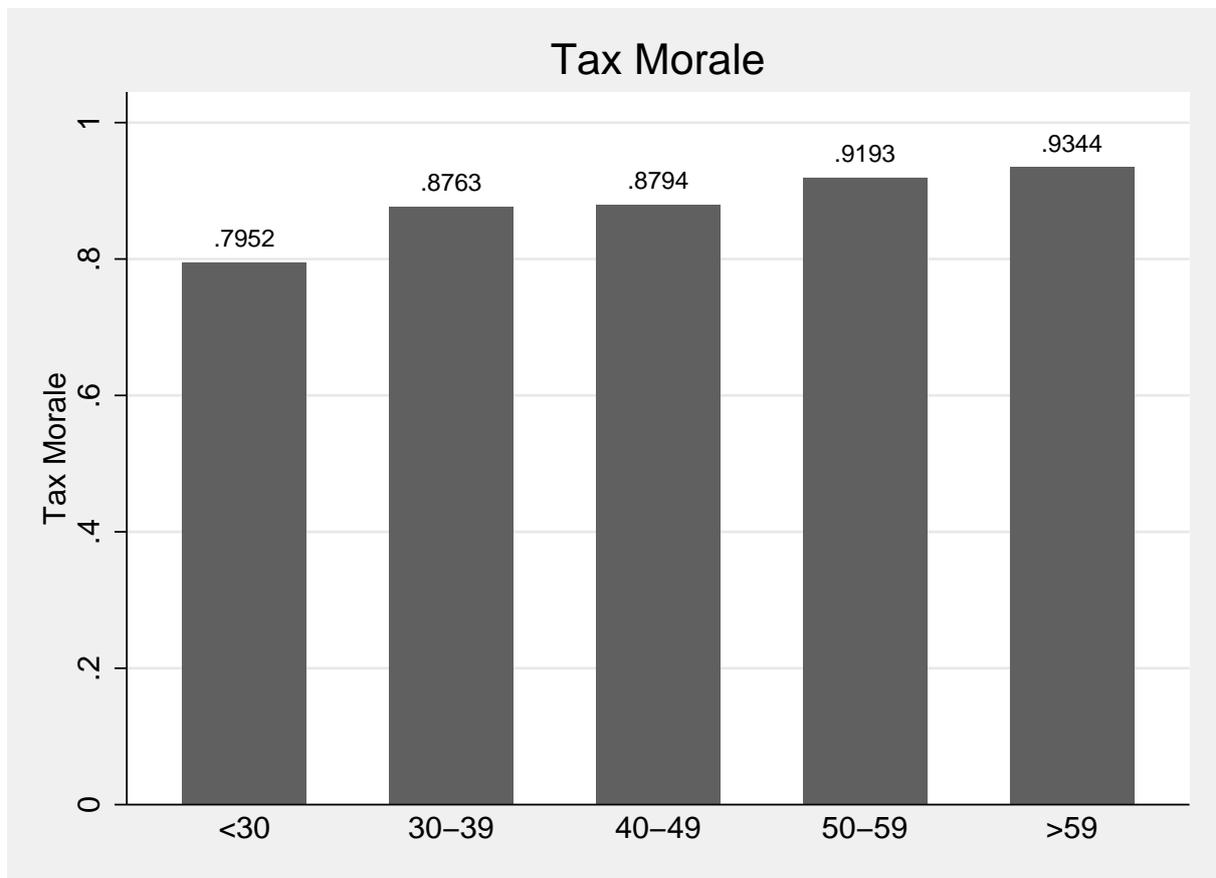
Notes: Randomization checks. The table shows the coefficients and robust standard errors (in parentheses) from a series of regressions of the form  $y_i = \beta Covariate_i + \epsilon_i$ , where *Covariate* is the respective variable that is listed. The dependent variables are dummies indicating the treatment groups. In Column (I),  $y_i$  is '1' if participant  $i$  is in the control group and '0' otherwise. In Column (II),  $y_i$  is '1' if participant  $i$  is in the social-norm group and '0' otherwise. In Column (III),  $y_i$  is '1' if participant  $i$  is in the reciprocity group and '0' otherwise. All covariates are defined as described in Table 1. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Figure 5: Tax Morale by Gender



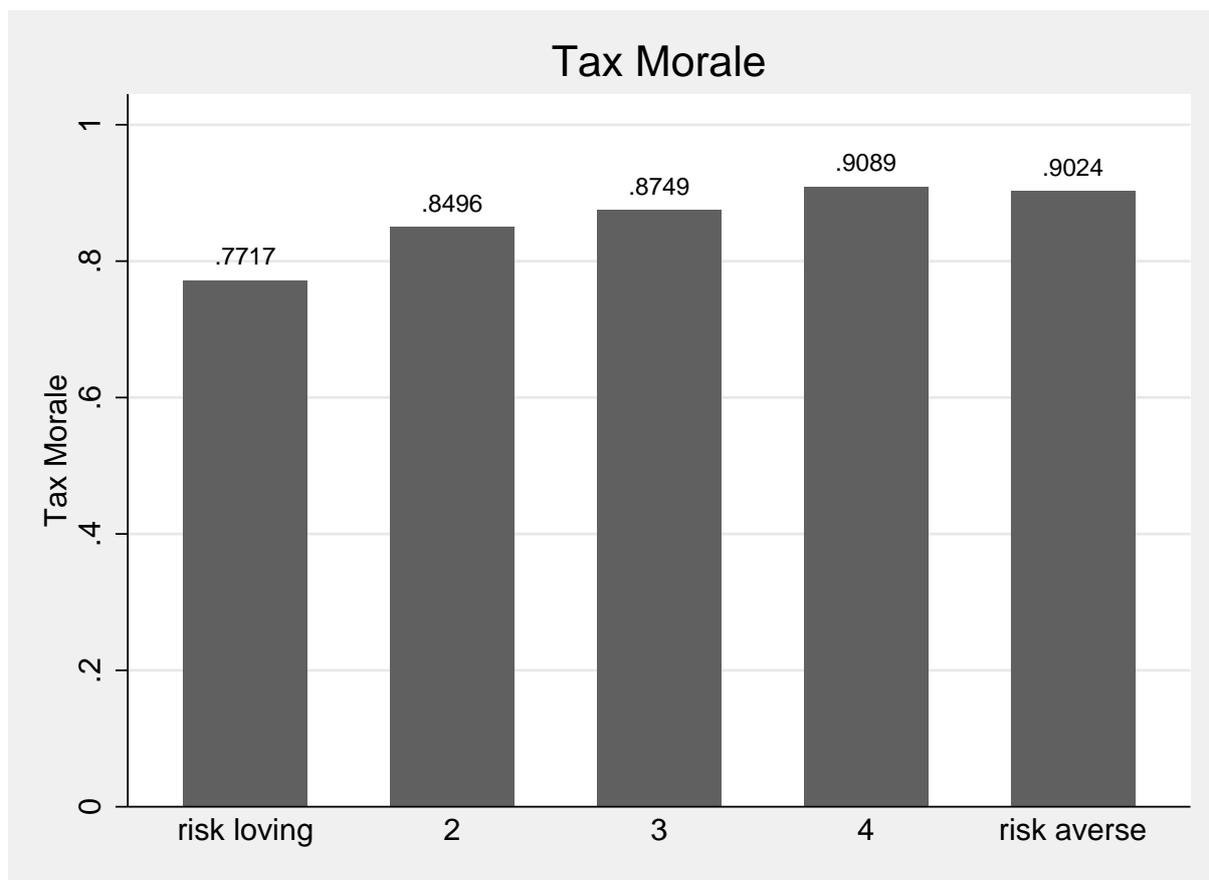
Notes: Average Tax Morale by gender. The outcome variable is survey-based tax morale as described in Section 2. Total number of observations is 3525. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Figure 6: Tax Morale by Age Categories



Notes: Average Tax Morale by age categories. The outcome variable is survey-based tax morale as described in Section 2. Total number of observations is 3525. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Figure 7: Tax Morale by Risk attitudes



Notes: Average Tax Morale by risk categories. The outcome variable is survey-based tax morale as described in Section 2. Total number of observations is 3525. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

Table 7: Anatomy of Tax Morale – incl. robust standard errors

Variable	Dep. var.: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
<u>Gender.</u> Reference category: <i>Female</i>					
Male	-0.041*** (0.011)	-0.042*** (0.012)	-0.038*** (0.012)	-0.035*** (0.013)	-0.035*** (0.013)
<u>Age.</u> Reference category: <i>&lt;30</i>					
30-39	0.063*** (0.022)	0.064** (0.025)	0.061** (0.025)	0.072** (0.029)	0.070** (0.029)
40-49	0.075*** (0.021)	0.089*** (0.025)	0.086*** (0.025)	0.076*** (0.029)	0.074** (0.029)
50-59	0.113*** (0.021)	0.117*** (0.024)	0.116*** (0.024)	0.097*** (0.028)	0.096*** (0.028)
>59	0.116*** (0.025)	0.111*** (0.029)	0.108*** (0.029)	0.096*** (0.032)	0.095*** (0.032)
<u>Marital status.</u> Reference category: <i>Not Married</i>					
Married	-0.001 (0.015)	0.007 (0.018)	0.004 (0.018)	-0.002 (0.021)	-0.001 (0.021)
<u>Size of household.</u> Reference category: <i>1</i>					
2	-0.006 (0.018)	-0.019 (0.021)	-0.018 (0.021)	-0.017 (0.024)	-0.018 (0.024)
3	-0.029 (0.022)	-0.052** (0.026)	-0.052** (0.026)	-0.038 (0.029)	-0.040 (0.029)
4	-0.029 (0.023)	-0.047* (0.027)	-0.047* (0.027)	-0.051 (0.032)	-0.052 (0.032)
>4	-0.032 (0.030)	-0.051 (0.035)	-0.058* (0.035)	-0.050 (0.040)	-0.052 (0.040)
<u>Unemployment.</u> Reference category: <i>Employed</i>					
Unemployed	-0.025 (0.038)	-0.031 (0.044)	-0.039 (0.045)	-0.045 (0.053)	-0.044 (0.053)
<u>Retirement Status.</u> Reference category: <i>Not Retired</i>					
Retired	0.012 (0.017)	0.030 (0.019)	0.029 (0.019)	0.035* (0.020)	0.034* (0.020)
<u>Education.</u> Reference category: <i>Low Education</i>					
2	0.057 (0.077)	0.065 (0.085)	0.059 (0.082)	-0.047 (0.102)	-0.045 (0.102)
3	0.099 (0.074)	0.085 (0.081)	0.078 (0.078)	-0.015 (0.097)	-0.012 (0.097)
4	0.131* (0.074)	0.126 (0.082)	0.117 (0.079)	0.014 (0.098)	0.019 (0.098)
High Educ	0.145** (0.074)	0.129 (0.081)	0.125 (0.079)	0.020 (0.097)	0.023 (0.098)
<u>East Germany.</u> Reference category: <i>West Germany</i>					
East Germany	-0.001 (0.013)	0.014 (0.014)	0.015 (0.014)	0.022 (0.016)	0.022 (0.016)
<u>Net household income.</u> Reference category: <i>Poor</i>					
2		0.017 (0.020)	0.014 (0.020)	0.037 (0.024)	0.037 (0.024)
3		0.045** (0.022)	0.044** (0.022)	0.062** (0.026)	0.062** (0.026)
rich		0.016 (0.029)	0.018 (0.028)	0.042 (0.032)	0.041 (0.032)
<u>Risk aversion.</u> Reference category: <i>Risk loving</i>					
2			0.058 (0.050)	0.060 (0.056)	0.060 (0.056)
3			0.089* (0.048)	0.093* (0.054)	0.093* (0.054)
4			0.118** (0.048)	0.109** (0.054)	0.109** (0.054)
risk averse			0.108** (0.048)	0.101* (0.054)	0.101* (0.054)
<u>Patience.</u> Reference category: <i>Not patient</i>					
2			0.041 (0.032)	0.016 (0.036)	0.017 (0.036)
3			0.029 (0.032)	0.015 (0.036)	0.015 (0.036)
4			0.033 (0.032)	0.029 (0.035)	0.028 (0.035)
patient			0.070** (0.031)	0.055 (0.035)	0.055 (0.035)
<u>Political preferences.</u> Reference category: <i>Conservative</i>					
social				0.007 (0.014)	0.007 (0.014)
right wing				-0.096*** (0.029)	-0.097*** (0.029)
left wing				-0.013 (0.025)	-0.012 (0.025)
<u>Experimental Treatment Group.</u> Reference category: <i>Control</i>					
Social Norm					-0.012 (0.016)
Reciprocity					0.014 (0.015)
constant	0.732*** (0.078)	0.715*** (0.085)	0.586*** (0.096)	0.700*** (0.119)	0.699*** (0.120)
N	3519	2881	2875	2236	2236
R2	0.036	0.039	0.049	0.054	0.056

Notes: Determinants of Tax Morale. Table identical to table 3, but including the presentation of robust standard errors (in parentheses).

Table 8: Robustness: Probit regressions

	Dep. Variable: Tax Morale		
	(I)	(II)	(III)
<u>Experimental Group.</u>	Reference category: <i>Control</i>		
Social Norm	-0.093 (0.067)	-0.115* (0.069)	-0.093 (0.067)
Reciprocity	0.079 (0.070)	0.077 (0.071)	0.079 (0.070)
p-val Norm vs Recipr.	0.012**	0.006***	0.012**
N	3525	3519	3519
Demographics	no	yes	no

Notes: The table presents the effects of the randomized treatment interventions on Tax Morale. Probit Regressions of Tax Morale on treatment dummies. The experimental groups are: Control group, Social-norm group and Reciprocity group. Control is omitted, implying that the effects are relative to the Control Group. Line *p-val Norm vs Recipr.* presents the p-values from chi-tests which compare if the regression coefficient for the Social-norm group is different from the the regression coefficient for the Reciprocity group. The treatment structure is augmented. To isolate the effect of the reciprocity treatment, one thus needs to compare the reciprocity treatment relative to the social-norm treatment. Columns (I) and (II) differ in the included sets of covariates. (I): no covariates, (II): gender, age, marital status, household size, employment status, retirement status, education and a dummy indicating East Germany. Column (III) does not include covariates, but is estimated based on the estimation sample in column (II). All variables are defined as described in Table 1. Robust Standard Errors in Parentheses. \* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).