Discussion Paper No. 17-045

Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment

Philipp Doerrenberg and Andreas Peichl

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Centre for European Economic Research

Discussion Paper No. 17-045

Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment

Philipp Doerrenberg and Andreas Peichl

Download this ZEW Discussion Paper from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/dp17045.pdf

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Tax morale and the role of social norms and reciprocity. Evidence from a randomized survey experiment*

Philipp Doerrenberg Andreas Peichl

November 3, 2017

Abstract

We present the first randomized survey experiment in the context of tax compliance to assess the role of social norms and reciprocity for intrinsic tax morale. We find that participants in a reciprocity treatment have significantly higher tax morale than those in a social-norm treatment. This suggests that a potential backfire effect of social norms is outweighed if the consequences of violating the social norm are made salient. We further document the anatomy of intrinsic motivations for tax compliance and present first evidence that previously found gender effects in tax morale are not driven by differences in risk preferences.

JEL Classification: H20, H32, H50, C93
Keywords: Tax compliance, Tax evasion, Intrinsic motivations, Tax morale, Social norms, Reciprocity

*__Doerrenberg__: ZEW Mannheim, CESifo and IZA. Email: doerrenberg@zew.de. **Peichl**: ifo Institute, University of Munich, CESifo, IZA and ZEW. Email: peichl@ifo.de. Sebastian Camarero Garcia, Christoph Feldhaus, Paul Hufe, Daniel Mueller, Carina Neisser, Jan Schmitz, Holger Stichnoth, Christian Traxler and Nils Wehrhoefer as well as various seminar participants provided helpful comments and suggestions. Eva Schoenwald and Felicitas von Dombois provided excellent research assistance. 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 penalties, audit probabilities and tax policy) but also by intrinsic non-pecuniary motives.[1] Following Luttmer and Singhal (2014), we use the term *tax morale* as an umbrella term for such intrinsic tax-compliance motives.[2] While tax morale is shaped by many factors (such as guilt, preferences for honesty, moral sentiments and cultural factors), social norms and reciprocity are often at the center of explaining its determinants. In this paper, we present the first randomized survey experiment in the context of tax compliance to assess the role of social norms and reciprocity for shaping tax morale. Moreover, we shed new light on the anatomy of tax morale.

Social norms of tax-compliance behavior particularly depend on the perception about the prevalence of tax evasion in society. Taxpayers are likely to 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. Evidence in this direction is presented by Paetzold and Winner (2016) who show that taxpayers evade more taxes after they change jobs to a firm where evasion is more common than in their previous firm. 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.[3]

Reciprocity in this context means that the motivation to comply depends on the (perceived) services and quality of the state/government which citizens receive in return for their tax payments (this is also related to the concept of conditional cooperation; see Frey and Meier 2004). Survey correlations and experimental evidence from the laboratory support the significance of reciprocity for tax compliance (Alm and Jackson 1993; Frey and Torgler 2007; Alm 2012; Lamberton et al. 2014). Evidence from randomized interventions outside the laboratory is somewhat mixed. While studies such as Blumenthal et al. (2001), Dwenger et al. (2016) and Castro and Scartascini (2015) do not find sig-

[1] Dwenger et al. (2016), for example, 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. Luttmer and Singhal (2014) provide a survey and summarize the role of non-pecuniary motives and intrinsic motivations on actual compliance in detail.

[2] 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.

[3] In line with this empirical evidence, 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). Lab-experimental work on tax compliance as well as evidence from other tax-related contexts also show that the behavior of other people affects individual behavior (Fortin et al. 2007; Chetty et al. 2013; Bohne and Nimczik 2017).

nificant effects of highlighting the services of the state, Bott et al. (2017) and Hallsworth et al. (2017) provide evidence in favor of reciprocity. As argued by Luttmer and Singhal (2014), null findings in this context should rather be attributed to the lack of power of the interventions (one sentence in a letter likely does not change views about the state which are formed over many years) than suggesting that reciprocity does not matter. Indeed, Besley et al. (2015) document that a strong intervention – the introduction of a tax that was widely perceived to be unfair – increased evasion; this suggests that government policy does affect compliance (holding enforcement constant) and that reciprocity matters.

We conduct a randomized survey experiment to shed new light on the drivers of tax morale. We particularly study the role of social norms and reciprocity as well as their interaction. The survey experiment is embedded in the German Internet Panel (GIP), a representative online survey in which we included a question measuring tax morale: 'How justifiable do you think it is to evade taxes if an easy opportunity to do so presents itself?'. This question is similar to the one used in the World Values Survey (WVS) which is widely used in the tax-compliance literature (e.g., Slemrod 2003, Alm and Torgler 2006). It 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, ensuring reliable replies. There is also evidence indicating that the replies to the corresponding question in the WVS are indeed linked with actual levels of tax evasion (Halla 2012). This question in the GIP is at the center of our paper.

Before responding to this question, participants were randomly assigned to three experimental groups. 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 that scientific studies estimate the tax gap in industrialized countries to be approximately 10%.[4] 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 that the government expenses for education in Germany could be increased by approximately 50% if the foregone revenue that is due to the tax gap would be spent on education. This variation adds a reciprocity component; 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. We label this group the 'reciprocity group'. All information provided

[4]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 the treatments were not deceptive (see Section 2 for a validation and justification of information and numbers referenced in the experimental treatments).

We find the following main results in our randomized survey experiment. 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 is not statistically significant in all regression specifications). This is in line with literature in other contexts (see below) and confirms 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. 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 (in line with literature findings in other contexts; see below) and that a backfire effect of social norms is outweighed if the consequences of the social norm are made salient.

The treatment effects are in the range of 2-3 percentage points. Given that only 11% of participants find tax evasion acceptable (i.e., the inverse of tax morale, $1-TaxMorale$) the effects are not only statistically but also economically significant. 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. In addition, our experimental manipulation consisted of only one or two additional sentences and was therefore fairly minor.

Apart from studying the effects of the randomized survey experiment, the survey also allows us to shed new light on the anatomy of tax morale. We confirm earlier findings on the (correlational) effects of gender and age on tax morale as 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 are driven by risk or patience, and additionally 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 on attitudes towards redistribution (Alesina et al. 2017), we find that participants with right-wing political attitudes have significantly lower tax morale.

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). As already discussed in footnote 1 above, there is evidence that tax morale exists and translates into actual tax-paying behavior. Such

findings motivate studies on tax morale and make them relevant.[5] 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, age or marital status. 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.

We further relate to a more recent literature that uses randomized variation to study tax compliance (e.g., Slemrod et al. 2001; Blumenthal et al. 2001; Kleven et al. 2011; Castro and Scartascini 2015; Pomeranz 2015; Dwenger et al. 2016; Boyer et al. 2016; Hallsworth et al. 2017; Bott et al. 2017; Drago et al. 2017). While these studies use randomized variation in the field and look at actually reported tax bases as outcome variables, we implement a randomized survey experiment with a focus on survey-reported tax morale. To our knowledge, this is the first paper to implement a randomized survey experiment in the context of tax compliance.[6]

Social norms and reciprocity have been studied extensively in the literature on public goods, where contributing to the public good can be interpreted as the equivalent choice to paying taxes honestly. This literature 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 shows 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. Section 3 presents the results with respect to the anatomy of tax morale. The results of the randomized survey experiment are presented and discussed in Section 4. Section 5 concludes the paper.

[5] 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."

[6] 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 redistributional 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. 2017).

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.[7] GIP data are collected online on a bi-monthly basis. Recruitment was conducted offline with face-to-face interviews, during which respondents were invited to the online panel. The survey is representative for the German population aged 16 to 75. 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.

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. The reference for GIP wave 14 is Blom et al. 2016). This wave went to the field in November 2014 and included 3575 participants. The data were released in 2016. We use wave 14 for our analysis and 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 a good opportunity to do so presents itself?*

Survey participants can reply to this question on a 6-point scale.[8] We use a binary version of the variable as the main outcome variable in our empirical analysis. The recoded 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

[7] 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/`. Examples of GIP-based papers include Kerschbamer and Müller (2017) and Müller and Renes (2017).

[8] The reply categories were: *very justifiable*, *justifiable*, *rather justifiable*, *rather not justifiable*, *not justifiable* and *not at all justifiable*. 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).

is more or less not justifiable) vs. low tax morale (evasion is more or less justifiable). While this dummy variable allows for an intuitive and simple interpretation of the results, we use the 6-point scale in a robustness check below.

The question is a slightly modified version of the commonly used question in the WVS.[9] As in the related WVS question, the 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 triggers reliable answers which reflect the true intrinsic motivation to pay taxes. 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). Despite the usual problems with such survey questions and surveys in general, the hypothetical character of the question along with this empirical evidence support the the view that tax morale can be measured with this question.

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. 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. 1178 out of 3532 overall 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*

[9]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'.*

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).[10] 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 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. 1177 out of 3532 overall participants were assigned to this social-norm group.

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 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. 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 probably agree that higher education expenses are better than lower education expenses. The treatment reminds participants

[10] 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).

that more compliance could yield better policy in return, and thus appeals to the reciprocity of participants.[11] 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. 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 EUR.[12] A tax gap of 10% then implies that the foregone revenues due to the tax gap amounts to about 69 Billion EUR (tax gap $= 0.1 = 69/(620 + 69)$). According to the Federal Statistical Office, the expenses for education in Germany in 2013 stood at 116 Billion Euro.[13] 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.

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

[11] 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.

[12] Source: `https://www.destatis.de/DE/Publikationen/Thematisch/FinanzenSteuern/Steuern/Steuerhaushalt/SteuerhaushaltJ2140400137004.pdf?__blob=publicationFile`

[13] Source: `https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2014/02/PD14_066_217.html`

(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 1500 and 3000 EUR) 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.

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.[14] 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. We take these findings as reliable evidence that our measure of risk attitudes is an appropriate measure for actual risk behavior.

We are not aware of a comparable study on the relation between self-reported patience and actual patience. The findings of Dohmen et al. (2011) on risk might provide an indication that self-reported measures of personal characteristics are good predictors of actual behavior – but this is only speculative. In our context of tax compliance it is particular interesting if some of the usual correlates of tax morale, such as gender, are confounded by omitting risk attitudes. This implies that a valid measure of risk is particularly valuable for our purposes. 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. Table 3 presents the results of randomization checks. Following the strategy in Alesina et al. (2017), we test balance across groups as follows:

[14]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).

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. As a result of this procedure, we have the results of 24 OLS regressions (one regression for each combination of 8 covariates and 3 outcome dummies).

Table 3 shows the p-values for these 24 regressions (robust standard errors). Overall, randomization worked very well. 7 out of 8 of the covariates cannot explain the treatment-group status. 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 addition, having some 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). We present 57 coefficients and find three significant effects which also corresponds to a share of about 5% and backs the assertion that randomization was successful. To circumvent any concerns with regard to randomization as good as possible, we show regression results with and without conditioning on covariates.

3 Anatomy of tax morale

Table 4 shows the estimates of a simple OLS regression of our outcome variable – tax morale – on different variables included in the survey. 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 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 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.[15] Specifi-

[15] These variables have a slightly lower number of non-missing observations, which is why we add them

cation (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,[16] 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.[17] It is thus an insight adding to previous literature that the previously found gender difference is not driven by risk aversion.[18]

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.[19] 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. The effects of marital status, household size and employment status on tax morale are not statistically significant. The effect of education looses significance as soon as income is added to the regression (specifications I and II).

subsequently.

[16]Doerrenberg and Peichl (2013) briefly summarize the survey literature with respect to drivers of tax morale.

[17]See the variable description in section 2 for a discussion on the reliability of our measure of risk 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.

[18]For illustration purposes, we show the unconditional means for men and women in Appendix Figure 5.

[19]The unconditional means for the different age groups are depicted in Figure 6 in the Appendix for illustration purposes.

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 3000 and 5000 EUR). Being in this income group increases tax morale by about 4 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., loosing 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 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 points higher than for risk-loving participants (category 1).[20] 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).[21]

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 9 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 5. 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

[20] Figure 7 in the Appendix shows unconditional levels of tax morale by risk attitudes.

[21] 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.

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.

The non-parametric findings are mirrored in the OLS regressions shown in Table 5. 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. In light of randomized variation, adding covariates does not change the treatment effects by large magnitudes, though the differences partly become more significant; see specifications (II) to (V). Adding demographics to the regression (II) slightly increases the difference between the control and social norm group (2.2% percentage points) and makes it significant on the 10% level. The difference between groups social norm and reciprocity comes with a p-value of 0.007 in this specification, and hence is significant on the 1% level.

Measures for risk and patience as additional covariates (III) leave the coefficients and significance levels unchanged. Specification (IV) adds household income as a covariate. This reduces the number of observations (from about 3500 to 2875) and vanishes the significance between the control and social-norm group. The difference between social-norm and reciprocity remains significant at the 1% level. Specification (V) additionally conditions on political preferences. This again reduces the number of observations (to about 2230) and leaves the differences between the control group and the treatment groups insignificant. The significance of the difference between the social-norm and reciprocity groups becomes weaker but remains at the 10% level.

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 all regression specifications). (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).[22] 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".[23]

The positive effect of adding the reciprocity component to the social-norm information (i.e., group reciprocity vs group social norm) 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 the citizens and that this contract is threatened through tax evasion. 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.[24]

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.

[22]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).

[23]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'."

[24]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.

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 6. 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. Also due to an effort of 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 in the table (significance stars based on robust standard errors and the usual levels of significance).

We overall do not find much heterogeneity of the treatments effects. 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. We find one heterogeneous effect of income; participants in the third income group (net household income between 3000 and 5000 EUR) 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 3000 EUR and '1' for more than 3000 EUR), 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. Results not reported in the regression table.)

In their randomized survey experiment on redistributive preferences, Alesina et al. (2017) find heterogeneous effects with respect to political preferences. Accordingly, we would expect that, for example, left-leaning respondents respond more to the reciprocity treatment in our experiment because they have a higher preference for government spend-

ing. We indeed find that left-wing participants respond slightly more positive to the reciprocity treatment (interaction coefficient of 0.076). This coefficient is just not significant though (p-value: 0.120) and should therefore not be given too much meaning. 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.

Robustness of main results. We use OLS regressions in all previous analyses. Table 7 presents the results from probit regressions with tax morale as the dependent variable (the table is equivalent to the main OLS regression table 5, 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, these estimates are only partly distinguishable from zero; in particular, the effect of social norms is statistically significant in specifications (II) and (III) 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.

While we use a simple tax morale dummy as the outcome in all preceding regressions, Table 8 presents ordered probit regressions using the 6-point scale version of the tax morale question as the outcome variable (this is the equivalent to Table 7, but using the 6-point scale outcome variable and ordered probit). The sign of the estimates is again comparable to all previous estimates; negative estimate for social norms and positive estimate for reciprocity. However, we lose statistical precision; the social-norm treatment is no longer significant in any of the specifications. The difference between the estimates for social norms and reciprocity are statistically significant only in specifications (IV) and (V) where we add a wide range of different covariates to the regression specification.

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 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.

Our second contribution is to conduct the first randomized survey experiment in the context of tax compliance. 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, relative to the social-norm treatment. 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.

References

Alesina, A., S. Stantcheva, and E. Teso (2017). Intergenerational mobility and support for redistribution. *American Economic Review*. forthcoming.

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 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.

Bertrand, M. (2011). New perspectives on gender. *Handbook of Labor Economics 4*(Part B), 1543 – 1590.

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 (2017). Learning dynamics in tax bunching at the kink:evidence from ecuador. mimeo available at `https://drive.google.com/file/d/0B2T3AGjlnKEyY0lMbFZ2cEVDSjQ/view`.

Bott, K. M., A. W. Cappelen, E. O. Sorensen, and B. Tungodden (2017). You've got mail: A randomised field experiment on tax evasion. NHH department of economics discussion paper SAM 10 2017.

Boyer, P. C., N. Dwenger, and J. Rincke (2016). Do norms on contribution behavior affect intrinsic motivation? field-experimental evidence from germany. *Journal of Public Economics 144*, 140 – 153.

Castro, L. and C. Scartascini (2015). Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization 116*, 65 – 82.

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.

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.

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.

Drago, F., F. Mengel, and C. Traxler (2017). Compliance behavior in networks: Evidence from a field experiment. `https://www.hertie-school.org/fileadmin/5_WhoWeAre/1_People_directory/Faculty_downloads/Traxler/Publications/DMT-Networks_most-recent.pdf`.

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.

Fehr, E. and S. Gaechter (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.

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.

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 behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14 – 31.

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.

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.

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.

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.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics 79*(3), 455 – 483.

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.

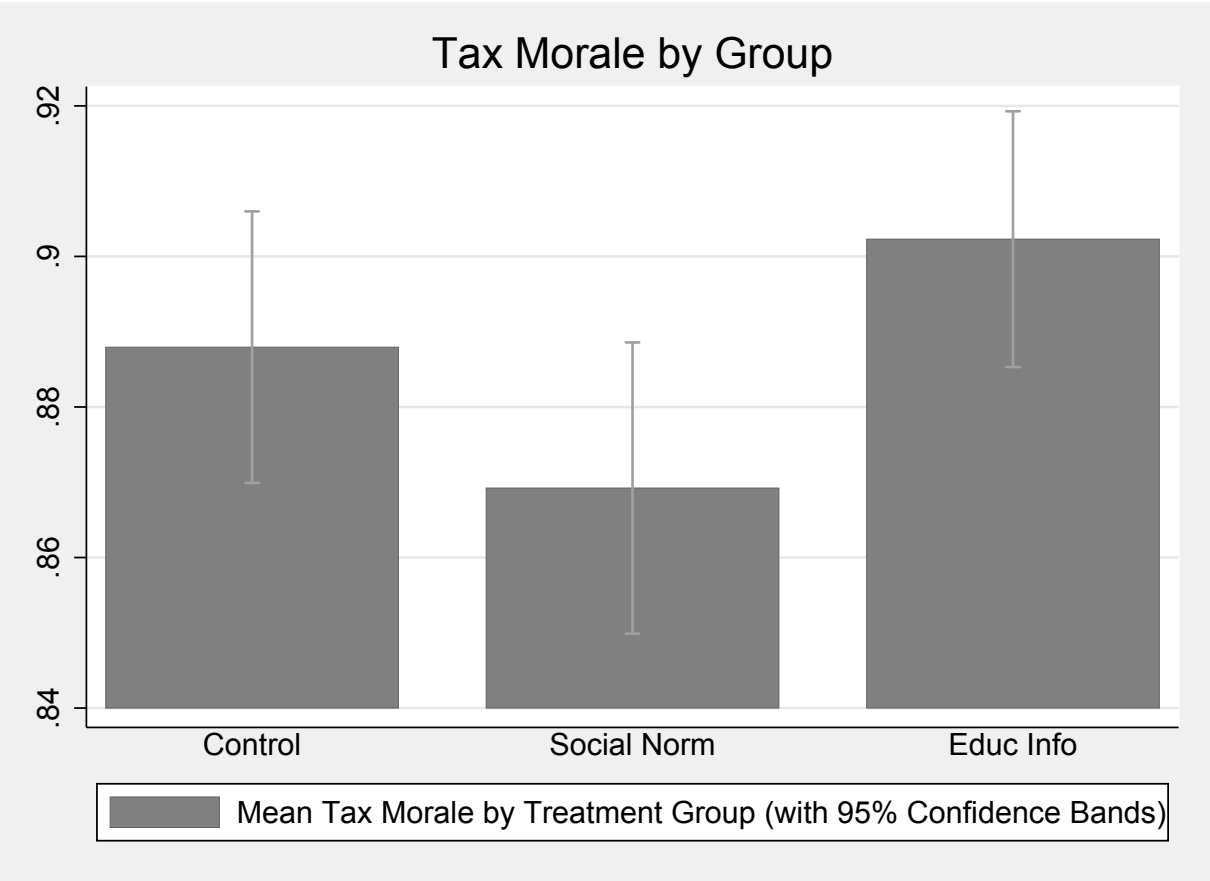
Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy 26*(1), 89 – 103.

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.

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. 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'	*CF14015*
Treatment status	(1): Control; (2) Social norm; (3) Reciprocity	*expCF1401*
Gender	(1): Male; (0): Female	*gender 14*
Age	(1): < 30; (2): 30-39; (3): 40-49; (4): 50-59; (5): > 59	*age cat 14*
Marital Status	(1): Married; (0): Not married	*marital status 14*
Household size	(1): 1; (2) 2; (3) 3; (4): 4; (5): > 4	*number hh members 14*
Employment status	(1): Unemployed; (0): Employed	*occupation 14*
Retirement Status	(1): Retired; (0): Not retired	*occupation 14*
Household Income (net)	(1): 0-1500 EUR; (2) 1500-3000 EUR; (3) 3000-5000 EUR; (4): > 5000	*AA1305x*
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	*ZE14074*
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	*AE14007*
Political preference	(1): Conservative; (2): Moderate left; (3): Right wing; (4) Left wing	*CE14140*
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	*educ school 14*

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	p10	p50	p90
Tax morale	3525	0.89	0.32	0.00	1.00	0.00	1.00	1.00
Control	3532	0.33	0.47	0.00	1.00	0.00	0.00	1.00
Social norm	3532	0.33	0.47	0.00	1.00	0.00	0.00	1.00
Reciprocity	3532	0.33	0.47	0.00	1.00	0.00	0.00	1.00
Gender	3574	0.49	0.50	0.00	1.00	0.00	0.00	1.00
Married	3575	0.58	0.49	0.00	1.00	0.00	1.00	1.00
Retired	3575	0.16	0.37	0.00	1.00	0.00	0.00	1.00
Unemployed	3575	0.03	0.16	0.00	1.00	0.00	0.00	0.00
Household size	3571	2.54	1.13	1.00	5.00	1.00	2.00	4.00
Age < 30	3573	0.18	0.38	0.00	1.00	0.00	0.00	1.00
Age 30-39	3573	0.16	0.37	0.00	1.00	0.00	0.00	1.00
Age 40-49	3573	0.19	0.40	0.00	1.00	0.00	0.00	1.00
Age 50-59	3573	0.23	0.42	0.00	1.00	0.00	0.00	1.00
Age > 60	3573	0.24	0.43	0.00	1.00	0.00	0.00	1.00
Conservative	2676	0.36	0.48	0.00	1.00	0.00	0.00	1.00
Moderate left	2676	0.44	0.50	0.00	1.00	0.00	0.00	1.00
Right wing	2676	0.11	0.31	0.00	1.00	0.00	0.00	1.00
Left wing	2676	0.09	0.29	0.00	1.00	0.00	0.00	0.00
Educ low	3574	0.01	0.09	0.00	1.00	0.00	0.00	0.00
Educ low-med	3574	0.08	0.27	0.00	1.00	0.00	0.00	0.00
Educ med	3574	0.52	0.50	0.00	1.00	0.00	1.00	1.00
Educ high-med	3574	0.15	0.35	0.00	1.00	0.00	0.00	1.00
Educ high	3574	0.24	0.43	0.00	1.00	0.00	0.00	1.00
Inc low	2919	0.15	0.36	0.00	1.00	0.00	0.00	1.00
Inc low-med	2919	0.40	0.49	0.00	1.00	0.00	0.00	1.00
Inc med	2919	0.35	0.48	0.00	1.00	0.00	0.00	1.00
Inc high	2919	0.10	0.30	0.00	1.00	0.00	0.00	1.00
Risk aversion	3517	3.67	1.12	1.00	5.00	2.00	4.00	5.00
Patience	3516	3.48	1.24	1.00	5.00	2.00	4.00	5.00

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: Randomization checks

	(I)	(II)	(III)
	Dep. var.: Treatment Group Indicator		
Variable	Control	Social Norm	Reciprocity
Gender. Reference category: *Female*			
Male	0.011	-0.010	0.009
	(0.016)	(0.016)	(0.016)
Age. Reference category: *<30*			
30-39	-0.026	-0.007	0.036
	(0.027)	(0.027)	(0.027)
40-49	-0.035	0.005	0.046*
	(0.026)	(0.026)	(0.026)
50-59	-0.012	0.002	0.022
	(0.025)	(0.025)	(0.024)
>59	-0.016	0.004	0.027
	(0.025)	(0.025)	(0.024)
Marital status. Reference category: *Not Married*			
Married	-0.021	0.017	0.011
	(0.016)	(0.016)	(0.016)
Size of household. Reference category: *1*			
2	0.010	-0.028	0.015
	(0.023)	(0.023)	(0.023)
3	-0.016	-0.022	0.039
	(0.027)	(0.027)	(0.027)
4	-0.026	-0.004	0.022
	(0.027)	(0.028)	(0.027)
>4	0.061	-0.058	-0.008
	(0.038)	(0.036)	(0.036)
Unemployment. Reference category: *Employed*			
Unemployed	-0.133***	0.105**	0.008
	(0.042)	(0.051)	(0.049)
Retirement Status. Reference category: *Not Retired*			
Retired	-0.033	0.017	0.021
	(0.021)	(0.022)	(0.022)
Education. Reference category: *Low Education*			
2	0.059	-0.056	-0.010
	(0.084)	(0.088)	(0.090)
3	0.064	-0.026	-0.052
	(0.080)	(0.085)	(0.086)
4	0.043	0.012	-0.066
	(0.082)	(0.087)	(0.088)
high educ	0.017	0.007	-0.034
	(0.081)	(0.086)	(0.087)
Net household income. Reference category: *Poor*			
2	-0.001	-0.009	0.013
	(0.026)	(0.026)	(0.026)
3	0.015	-0.022	0.007
	(0.027)	(0.027)	(0.027)
rich	-0.048	0.002	0.050
	(0.034)	(0.036)	(0.036)

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).

Table 4: Anatomy of Tax Morale

Variable	Dep. var.: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
Gender. Reference category: *Female*					
Male	-0.041***	-0.042***	-0.038***	-0.035***	-0.035***
	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)
Age. Reference category: *<30*					
30-39	0.063***	0.065**	0.061**	0.073**	0.072**
	(0.022)	(0.025)	(0.025)	(0.029)	(0.029)
40-49	0.075***	0.089***	0.086***	0.076***	0.074**
	(0.021)	(0.025)	(0.025)	(0.029)	(0.029)
50-59	0.113***	0.119***	0.118***	0.098***	0.097***
	(0.021)	(0.024)	(0.024)	(0.028)	(0.028)
>59	0.116***	0.112***	0.109***	0.097***	0.096***
	(0.025)	(0.029)	(0.029)	(0.032)	(0.032)
Marital status. Reference category: *Not Married*					
Married	-0.001	0.006	0.004	-0.002	-0.001
	(0.015)	(0.018)	(0.018)	(0.021)	(0.021)
Size of household. Reference category: *1*					
2	-0.006	-0.017	-0.016	-0.014	-0.016
	(0.018)	(0.021)	(0.021)	(0.024)	(0.024)
3	-0.029	-0.051**	-0.050*	-0.036	-0.038
	(0.022)	(0.026)	(0.026)	(0.029)	(0.029)
4	-0.029	-0.045	-0.046*	-0.049	-0.050
	(0.023)	(0.027)	(0.027)	(0.032)	(0.032)
>4	-0.032	-0.050	-0.056	-0.048	-0.049
	(0.030)	(0.035)	(0.035)	(0.040)	(0.040)
Unemployment. Reference category: *Employed*					
Unemployed	-0.025	-0.031	-0.039	-0.045	-0.045
	(0.038)	(0.045)	(0.045)	(0.053)	(0.053)
Retirement Status. Reference category: *Not Retired*					
Retired	0.012	0.030	0.029	0.035*	0.034*
	(0.017)	(0.019)	(0.019)	(0.020)	(0.020)
Education. Reference category: *Low Education*					
2	0.057	0.064	0.058	-0.048	-0.046
	(0.077)	(0.084)	(0.082)	(0.102)	(0.102)
3	0.099	0.085	0.079	-0.014	-0.011
	(0.074)	(0.081)	(0.078)	(0.097)	(0.097)
4	0.131*	0.126	0.118	0.014	0.019
	(0.074)	(0.082)	(0.079)	(0.098)	(0.098)
high educ	0.145**	0.131	0.127	0.023	0.026
	(0.074)	(0.081)	(0.079)	(0.098)	(0.098)
Net household income. Reference category: *Poor*					
2		0.015	0.012	0.035	0.035
		(0.020)	(0.020)	(0.023)	(0.023)
3		0.042*	0.041*	0.057**	0.057**
		(0.022)	(0.022)	(0.026)	(0.026)
rich		0.012	0.014	0.036	0.035
		(0.028)	(0.028)	(0.031)	(0.032)
Risk aversion. Reference category: *Risk loving*					
2			0.057	0.059	0.059
			(0.050)	(0.056)	(0.056)
3			0.089*	0.093*	0.092*
			(0.048)	(0.054)	(0.054)
4			0.117**	0.109**	0.108**
			(0.048)	(0.054)	(0.054)
risk averse			0.107**	0.101*	0.101*
			(0.048)	(0.054)	(0.054)
Patience. Reference category: *Not patient*					
2			0.041	0.016	0.017
			(0.032)	(0.036)	(0.036)
3			0.030	0.016	0.016
			(0.032)	(0.036)	(0.036)
4			0.034	0.029	0.029
			(0.032)	(0.035)	(0.035)
patient			0.070**	0.055	0.055
			(0.031)	(0.035)	(0.035)
Political preferences. Reference category: *Conservative*					
social				0.008	0.007
				(0.014)	(0.014)
right wing				-0.094***	-0.094***
				(0.028)	(0.029)
left wing				-0.007	-0.006
				(0.024)	(0.024)
Experimental Treatment Group. Reference category: *Control*					
Social Norm					-0.011
					(0.016)
Educ Info					0.015
					(0.015)
constant	0.732***	0.717***	0.589***	0.703***	0.701***
	(0.078)	(0.085)	(0.096)	(0.119)	(0.121)
N	3519	2881	2875	2236	2236
R2	0.036	0.039	0.049	0.054	0.055

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. 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: Effect of experimental intervention on tax morale

	Dependent variable: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
Experimental Group. Reference category: *Control*					
Social Norm	-0.019	-0.022*	-0.022*	-0.018	-0.011
	(0.013)	(0.013)	(0.013)	(0.015)	(0.016)
Reciprocity	0.014	0.013	0.013	0.021	0.015
	(0.013)	(0.012)	(0.012)	(0.014)	(0.015)
constant	0.888***	0.737***	0.602***	0.588***	0.701***
	(0.009)	(0.078)	(0.089)	(0.097)	(0.121)
p-val Norm vs Recipr.	0.012**	0.007***	0.006***	0.005***	0.096*
N	3525	3519	3498	2875	2236
R2	0.002	0.038	0.048	0.051	0.055
Demographics	no	yes	yes	yes	yes
Risk & Patience	no	no	yes	yes	yes
Household Income	no	no	no	yes	yes
Political Preference	no	no	no	no	yes

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. Columns (I)-(V) differ in the included sets of covariates. (I): no covariates, (II): gender, age, marital status, household size, employment status, retirement status, and education, (III): (II) plus risk aversion and patience, (IV): (III) plus net household income, (V): (IV) plus political preferences. 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 6: Heterogeneous effects of experimental interventions

	Dep. var.: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
Age. 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				
Marital status. Reference category: *Not married*					
Norm × Married		0.050*			
Recipr. × Married		0.015			
Size of Household.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		
Political preferences.Reference category: *Conservative*					
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	
Income Group.Reference category: *1*					
Norm × inc gr 2					-0.002
Norm × inc gr 3					0.107**
Norm × inc gr 4					0.025
Recipr. × inc gr 2					-0.070
Recipr. × inc gr 3					0.014
Recipr. × inc gr 4					-0.037
N	3523	3525	3522	2654	2886
R2	0.028	0.006	0.012	0.014	0.010

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).

Table 7: Robustness: Probit regressions

	Dependent variable: Tax Morale				
	(I)	(II)	(III)	(IV)	(V)
Experimental Group. Reference category: *Control*					
Social Norm	-0.093	-0.115*	-0.124*	-0.095	-0.067
	(0.067)	(0.069)	(0.069)	(0.075)	(0.088)
Reciprocity	0.079	0.077	0.079	0.129	0.098
	(0.070)	(0.071)	(0.072)	(0.079)	(0.091)
constant	-1.216***	-0.626**	-0.051	0.002	-0.437
	(0.048)	(0.278)	(0.320)	(0.345)	(0.518)
p-val Norm vs Recipr.	0.012**	0.006***	0.004***	0.004***	0.066*
N	3525	3519	3498	2875	2236
Demographics	no	yes	yes	yes	yes
Risk & Patience	no	no	yes	yes	yes
Household Income	no	no	no	yes	yes
Political Preference	no	no	no	no	yes

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. Columns (I)-(V) differ in the included sets of covariates. (I): no covariates, (II): gender, age, marital status, household size, employment status, retirement status, and education, (III): (II) plus risk aversion and patience, (IV): (III) plus net household income, (V): (IV) plus political preferences. All variables are defined as described in Table 1; exception is tax morale which is measured on 6-point scale rather than a dummy. 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 8: Robustness: Tax morale measured on 6pt scale, Ordered Probit regressions

	Dependent variable: Tax Morale (6pt scale)				
	(I)	(II)	(III)	(IV)	(V)
Experimental Group. Reference category: *Control*					
Social Norm	-0.027	-0.039	-0.037	-0.053	-0.017
	(0.046)	(0.046)	(0.046)	(0.051)	(0.058)
Reciprocity	0.014	0.010	0.015	0.042	0.080
	(0.045)	(0.045)	(0.045)	(0.050)	(0.057)
constant	0.158***	0.627***	1.040***	1.044***	1.051***
	(0.034)	(0.231)	(0.265)	(0.284)	(0.379)
p-val Norm vs Recipr.	0.365	0.281	0.251	0.057*	0.088*
N	3525	3519	3498	2875	2236
Demographics	no	yes	yes	yes	yes
Risk & Patience	no	no	yes	yes	yes
Household Income	no	no	no	yes	yes
Political Preference	no	no	no	no	yes

Notes: The table presents the effects of the randomized treatment interventions on Tax Morale. Ordered Probit Regressions of Tax Morale on treatment dummies. Tax morale is measured on a 6-point scale using all reply categories from the survey. 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 regression coefficient for the Reciprocity group. Columns (I)-(V) differ in the included sets of covariates. (I): no covariates, (II): gender, age, marital status, household size, employment status, retirement status, and education, (III): (II) plus risk aversion and patience, (IV): (III) plus net household income, (V): (IV) plus political preferences. 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).

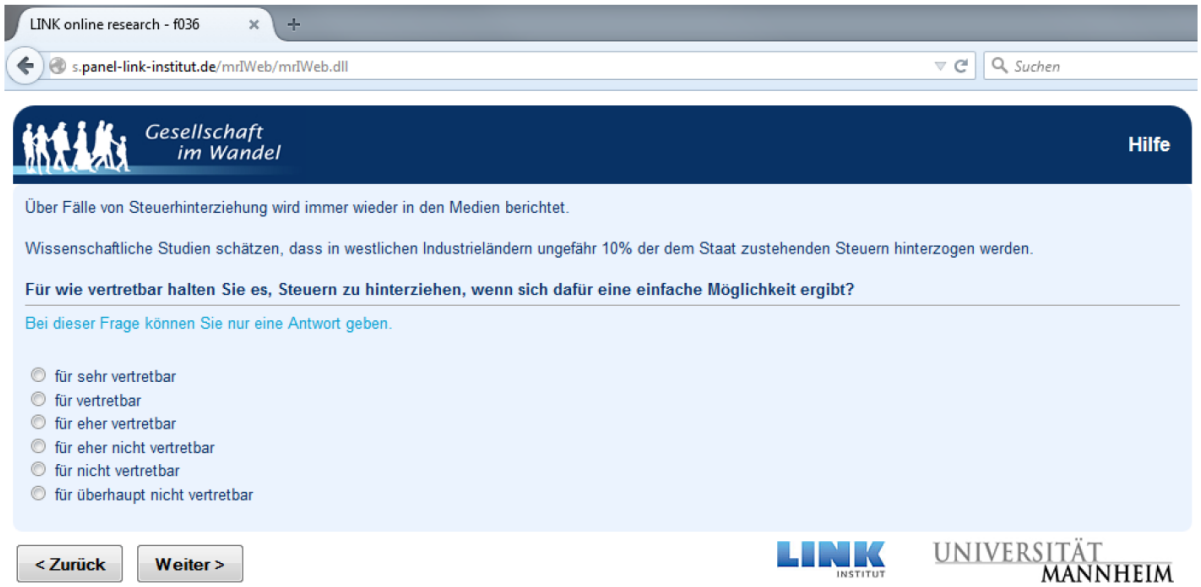
Appendix

Additional Figures

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/` and (Blom et al. 2016).

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

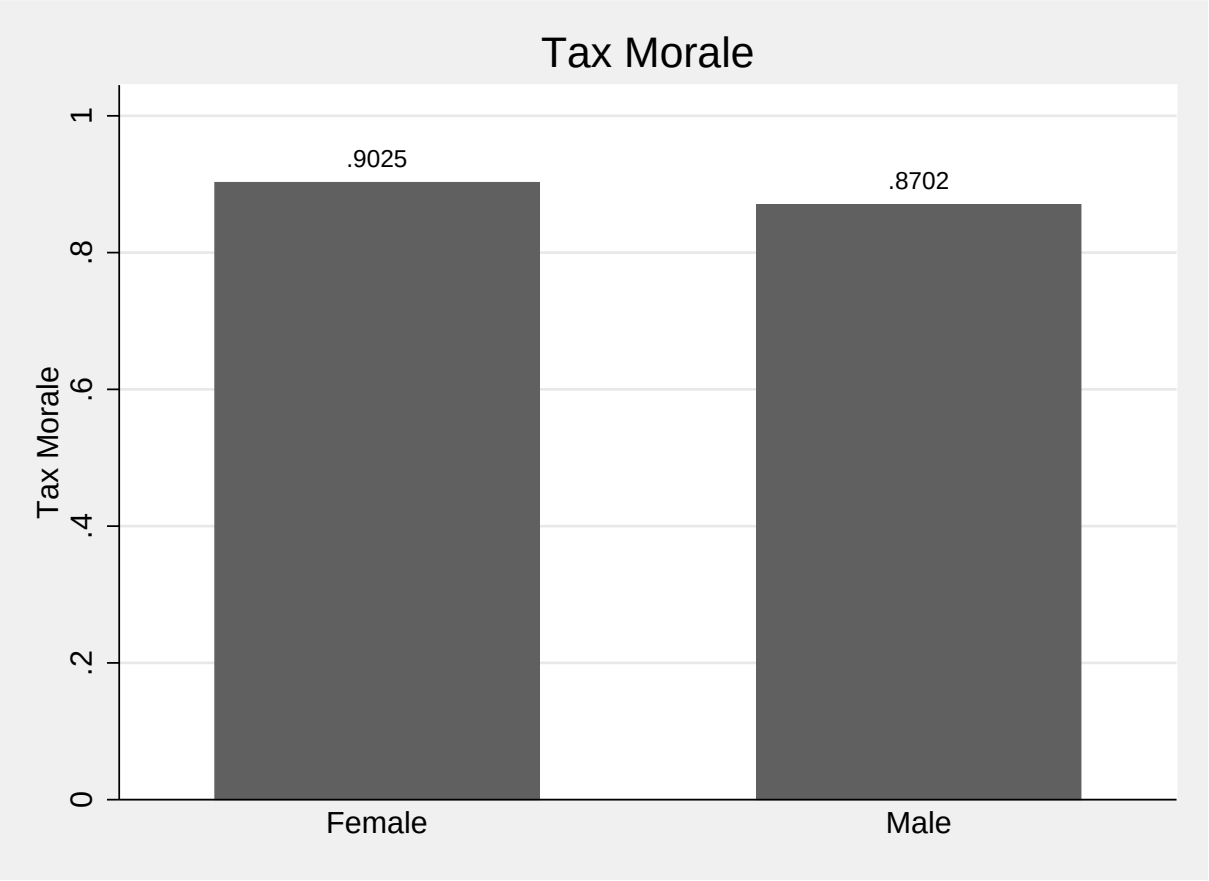


Notes: The figure shows a screenshot of the survey question which is used to measure tax morale. This screenshot shows the screen which is shown to participants in the experimental *social-norm group*. See section 2 for a description of the survey and the randomized survey experiment. German Internet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

Figure 4: Screenshot of survey experiment: reciprocity group

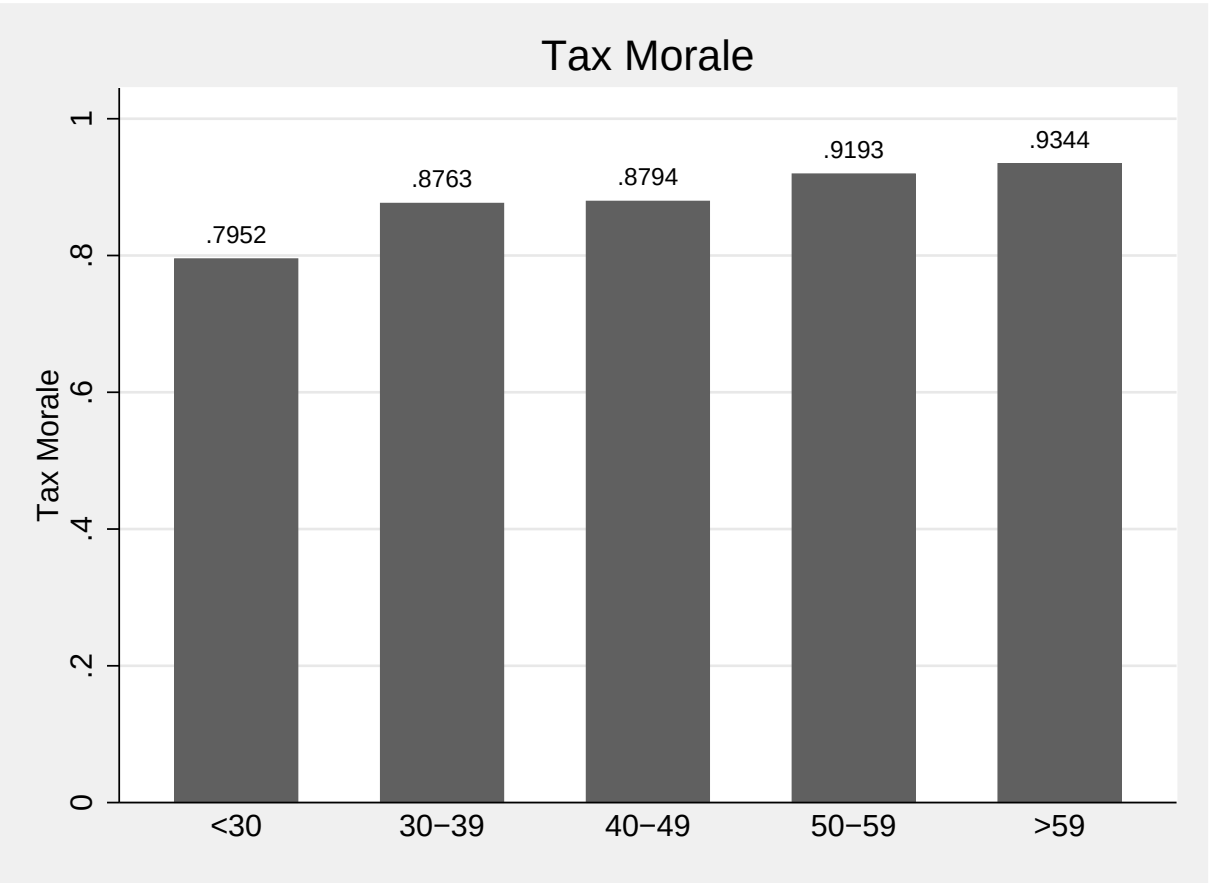
Notes: The figure shows a screenshot of the survey question which is used to measure tax morale. This screenshot shows the screen which is shown to participants in the experimental *reciprocity group*. See section 2 for a description of the survey and the randomized survey experiment. German Internet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

Figure 5: Tax Morale by Gender



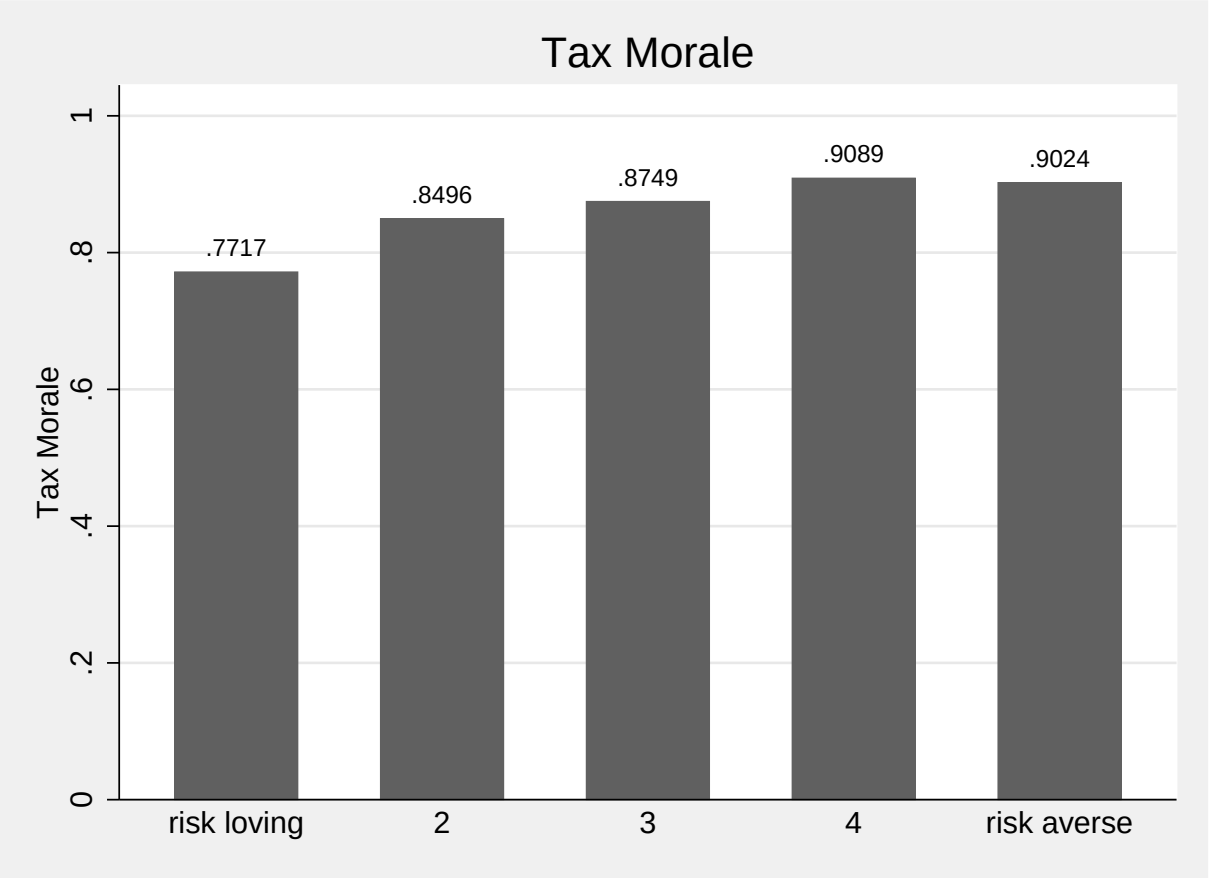
Notes: Average Tax Morale by gender. The outcome variable is survey-based tax morale as described in Section 2. Total number of observations is 3525. Data come from German Internet Panel (GIP) wave 14 (Blom et al. 2016).

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).