Accepted Manuscript

Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker

Uwe Dulleck, Jonas Fooken, Cameron Newton, Andrea Ristl, Markus Schaffner, Benno Torgler

PII: S0047-2727(15)00211-X
DOI: doi: 10.1016/j.jpubeco.2015.12.007
Reference: PUBEC 3635

To appear in: Journal of Public Economics

Received date: 13 November 2012
Revised date: 16 December 2015
Accepted date: 22 December 2015

Please cite this article as: Dulleck, Uwe, Fooken, Jonas, Newton, Cameron, Ristl, Andrea, Schaffner, Markus, Torgler, Benno, Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker, Journal of Public Economics (2016), doi: 10.1016/j.jpubeco.2015.12.007

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Tax Compliance and Psychic Costs: Behavioral Experimental Evidence Using a Physiological Marker

Uwe Dulleck¹,², Jonas Fooker¹, Cameron Newton³, Andrea Ristl¹, Markus Schaffner¹, and Benno Torgler¹,⁵

Abstract
Although paying taxes is a key element of a well-functioning society, there is still limited understanding as to why people actually pay their taxes. Models emphasizing that taxpayers make strategic, financially motivated compliance decisions seemingly assume an overly restrictive view of human nature. Law abidance may be more accurately explained by social norms, a concept that has gained growing importance as research attempts to understand the tax compliance puzzle. This study analyzes the influence of psychic stress generated by the possibility of breaking social norms in the tax compliance context. We measure psychic stress using heart rate variability (HRV), which captures the psychobiological or neural equivalents of psychic stress that may arise from the contemplation of real or imagined actions, producing immediate physiologic discomfort. The results of our laboratory experiments provide empirical evidence of a positive correlation between psychic stress and tax compliance, thus underscoring the importance of moral sentiments for tax compliance. We also identify three distinct types of individuals who differ in their levels of psychic stress, tax morale, and tax compliance.

JEL Classification: H26, H41, K42, D31, D63, C91
Keywords: tax compliance, psychic costs, stress, tax morale, cooperation, heart rate variability, biomarkers, experiment, heterogeneous individuals

¹Queensland Behavioural Economics Group (QuBE), School of Economics and Finance, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. ²Research School of Economics, Australian National University, Canberra, ACT, Australia. ³School of Management, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. ⁴AUTONOM TALENT® Consulting GmbH, Mariahilfer Straße 54/15, 1070 Vienna, Austria. ⁵CREMA – Center for Research in Economics, Management and the Arts, Gellertstrasse 18, CH-4052 Basel, Switzerland. This work was supported by the Australian Research Council (ARC), Linkage Grant LP0884074 and Future Fellowship FT11010463. We thank Brian Erard, Jonathan Feinstein, Richard Jefferson, and the participants of several seminars (University of New South Wales, Deakin University, University of Technology, Sydney, and Uppsala University), the shadow2011 conference in Muenster, the Workshop on Behavioural Economics: Science, Philosophy, and Policy-Making (2013) in Trento, Italy, the Taxation, Social Norms and Compliance (2014) conference in Nuremberg, Germany, the editor Kai A. Konrad and three anonymous referees for very helpful comments. Uwe Dulleck would like to dedicate his share of this work to the memory of Jürgen Wolters (1940-2015), his first mentor and teacher of econometrics.

* Correspondence should be addressed to: Benno Torgler (benno.torgler@qut.edu.au).
1. Introduction

Above the entrance to the U.S. Internal Revenue Service (IRS) are inscribed the following words from Oliver Wendell Holmes: “Taxes are what we pay for a civilized society.” Taxes are the fuel of civilizations, and few civilizations have failed to impose them (Adams, 1993). Yet despite the crucial importance of taxation, our understanding of why people pay taxes remains limited (see, e.g., Slemrod, 1992; Torgler, 2007; Alm, et al., 2010; Konrad and Qari, 2012). In this article we provide physiological evidence of an intrinsic explanation for tax compliance.

The traditional approach is to focus on the effects of deterrence policies, a major component of the economics-of-crime approach (Becker, 1968; Allingham and Sandmo, 1972). Although this approach has produced many useful insights, it seems too narrow a framework for fully understanding tax compliance if it assumes a rational individual who maximizes the expected utility of the tax evasion gamble, weighing the benefits of successful cheating against the prospect of detection and punishment. Using this framework, tax evasion is negatively related to the probability of detection and the degree of punishment. Yet, in many countries, the level of deterrence is too low to explain the high degree of tax compliance, and even in the least compliant nations, an approach based purely on this standard economic theory can only explain observed levels of tax evasion by assuming unrealistic preferences like extreme risk aversion (Alm et al., 2010). This implies that tax compliance decisions must be influenced by motivations not fully captured by the basic economics-of-crime approach.
The inability of the economics-of-crime model to explain compliance is probably related to a variety of factors, including social norms and social and cognitive dissonance, as theorized in the psychological literature and conceptualized on both social and individual levels (Asch, 1952; Festinger, 1957). It has even been argued (e.g., Graetz and Wilde, 1985; Alm, McClelland et al., 1992) that social norms act as an alternative to law enforcement in motivating individual behavior. Investigating these norms may therefore help resolve the puzzle of (high) tax compliance. To influence tax payment decisions, social norms must be internalized; for example, through moral sentiments that create psychological costs or stress in case of norm violation. The (intended) violation of social norms triggers internal sanctioning (e.g., moral sentiments like guilt, remorse, or shame), augmenting effects of legal and social sanctions such as prosecution, gossip, or ostracism in the case of detection (Polinsky and Shavell, 2000). Internal sanctions could function by creating increased psychic stress (potentially varying across taxpayers according to how much the social norm is internalized).

We study the role of psychic stress in tax compliance decisions using an experiment that requires participants to make compliance choices while their heart rate variability (HRV) is recorded. HRV is an established correlate to psychic strain (Dishman et al. 2000), and hence a physiological marker for psychic stress, capturing psychobiological or neural equivalents of psychic strain that arise from the contemplation of real or imagined actions (Green and Paxton 2009). It thus allows exploration of immediate physiological reactions represented by changes in the heart rate during participants’ decision making. Our use of neuroscientific tools to examine the possible impact of psychic stress on tax compliance addresses the valid criticism that past research on moral sentiments has failed to reliably measure psychic stress (Andreoni et al., 1998). Our work thus has the potential to contribute to a better
biological microfoundation for compliance behavior. In general, inclusion of data from brain systems, heart rate, skin resistance, genes, neurons, and neurotransmitters can reveal otherwise unobservable aspects of the individual decision-making process. For example, functional magnetic resonance imaging (fMRI) evidence suggests that mandatory (tax-like) transfers to charity lead to neural activation in areas linked to reward processing (Harbaugh et al., 2007).

Noninvasive procedures like HRV are particularly attractive for understanding the biological bases of human decisions in a laboratory setting (Dulleck et al., 2011). Coricelli et al. (2010), for instance, use skin conductance responses (SCR) to explore the link between physiological measures and tax evasion. They find that SCR is correlated with self-reported emotional arousal and hedonic valence. Specifically, they link an increased SCR before decision making (i.e., higher anticipated and anticipatory emotional arousal) to a greater degree of tax evasion. Such a result is consistent with the idea that detected evasion creates shame, while in our framework the intention to evade taxes induces guilt, which consequently reduces the utility of noncompliance and thus the amount of tax evasion.²

We chose the HRV measurement technique because it is nonintrusive and requires only compact equipment, which allows us to design a complex environment (i.e., a large variety of settings and interpersonal contexts). Additionally, HRV permits observation of a larger group of participants and interactions in the laboratory compared to alternatives such as fMRI. Our experiment can therefore include a larger sample than is typically possible with fMRI studies.

² Coricelli et al. (2010) also observe that negative feelings are related to a higher fine, and learning that photographs of participants would be disseminated increases this effect. This supports their explanation that tax compliance is driven by shame, which is increased with greater public punishment. They also find that avoiding an audit generates positive feelings, possibly out of relief or the joy of higher earnings, which also hints at the importance of external sanctioning mechanisms.
2. Tax compliance and psychic stress

Erard and Feinstein (1994) describe the role of moral sentiments in tax compliance, using a theoretical model that outlines how these sentiments could influence decisions. Because our approach is best understood in the light of their theory, we summarize their framework as it identifies two central moral sentiments connected to tax evasion: shame and guilt. Both have an influence on tax compliance as guilt and shame arise from the contemplation of imagined actions. However, guilt is seen to be more anticipatory (Turner and Stets, 2005). Erard and Feinstein relate shame to the imagination of being caught and punished. In contrast, guilt is felt independently of detection; it is exclusively an internal sanctioning mechanism that arises when the individual contemplates violating an internal moral norm. In both cases, Erard and Feinstein’s model suggest that it is the mere contemplation of an incorrect declaration that generates psychic costs, which produce a desire to reduce this stress by reducing evasion.

We are not able to measure psychic costs directly but they are related to pain or negative emotions. The actual pain experienced while contemplating cheating behavior serves as a survival signal to demonstrate that corrective behavior is required (Churchland 2011). In our case the corrective behavior would be to reduce tax evasion. Thus, experienced pain serves as a *homeostatic emotion* reflecting an adverse condition in the body that prompts a behavioral reaction (such as being more compliant) to remove or reduce the pain (Craig 2003). Thus, by measuring such a body reaction with a physiological HRV marker we can determine whether pain or stress is related to the actual final decision (level of tax compliance). According to Erard and Feinstein’s (1994) model, psychic cost has a positive influence on tax compliance. In other words, if their model is correct we would observe that a higher
level of psychic stress measured during the decision making process (when individuals contemplate what to do) would be correlated with a higher level of tax compliance (final decision). The theoretical framework of moral sentiments in tax compliance decisions is rarely tested; however, survey evidence suggests that the anticipated psychic stress associated with tax evasion can serve as a much stronger compliance enforcement mechanism than the perceived threat of legal sanctions (Grasmick and Bursik, 1990).

The following introduces Erard and Feinstein’s theoretical model in order to provide an understanding of how psychic costs can be introduced into the standard setting. As a starting point, we introduce the common approach taken in the tax compliance literature. Agents choose an optimal amount of income to report \( D_i \) taking into account the probability that they will be audited and penalized (if audited and underreporting is discovered). In our experimental design, we assume a proportional tax \( t \) and an exogenous audit probability \( p \) (i.e., we do not assume endogenous audit selection rules like dependence on the amount reported). If the individual is audited (which happens with probability \( p \)), then her true income \( Y_i \) will be determined and the difference between true tax liability and declared income \( t(Y_i - D_i) \) will be taxed with a fine multiplier \( f(t(Y_i-D_i)) \). For example, for \( f=2 \), participants pay a fine equal to the amount they evaded on top of the outstanding tax. The following expected utility model describes the decision for a risk-neutral taxpayer:

\[
E[u(D_i)] = p[Y_i - tD_i - f t(Y_i - D_i)] + (1 - p)[Y_i - tD_i]
\] (1)

Erard and Feinstein change the expected utility framework, using a nonnegative guilt and shame parameter. In their model they modify the previous expression by adding guilt in the „not audited“ state and shame in the „audited“ state,
as the subject contemplating the embarrassment of an audit is subsequently subjected to embarrassment of an audit. Both psychic cost terms have the same structure. We modify their model by using only one term that we call psychic stress, measuring the negative emotions when contemplating tax evasion and therefore before knowing whether or not they have been caught. Thus, psychic stress affects the entire outcome term in (1).

\[
\left[ 1 + g_i \frac{1}{t(Y_i-D_i)/Y_i} \right] \left\{ p[Y_i - tD_i - ft(Y_i - D_i)] + (1 - p)[Y_i - tD_i] \right\} \tag{2}
\]

This equation redefines the utility state so that the denominator of the new expression becomes larger than one when a taxpayer *contemplates* underpaying and feels guilty about it, diminishing his or her overall utility level. Erard and Feinstein assume that taxpayer psychic costs are proportional to the ratio of underpaying \( t(Y_i-D_i) \) to the true income \( Y_i \). Parameter \( g_i \) also takes into account that the psychic stress experienced may vary across taxpayers.

The best way to determine whether or not we observe a positive correlation between tax compliance and psychic stress is to conduct a laboratory experiment that allows measurement of a physiological reaction during the decision process within a controlled environment. Thus, contrary to common experimental practice, rather than using the laboratory experiment to explore differences between a control and a treatment group, the primary aim of our application is to test whether the relationship between psychic stress and tax compliance persists when applying different settings used in past tax compliance experiments.
The main decision process starts when the income declaration screen appears and ends when the participants press the okay button (see Figure A1). The pressing of the okay button is described as point “0” in Figure 1. The average time it took a participant to make a decision once the decision screen came up was 15.61s, the standard deviation was 15.72s. 90% of the observations fall in the interval [3.41s..43.58s]. The figure also provides information on the waiting time and the audit phase after the decision. If the emotional reaction is due to the outcome itself or to the anxiety over the resolution of the audit uncertainty, we would expect to observe the maximum emotional activation in the interval after the outcome itself (waiting phase, see Figure 1). This was not the case. The strongest physiological reaction or stress response is observed approximately 17 seconds before the participants click the „OK“ button, which indicates that something is happening during the decision making process when subjects are considering which actions to take. Thus, this could indicate that contemplating mechanisms trigger stress responses although we cannot actually be sure that participants were contemplating tax evasion. Therefore, we can only report correlations (albeit with a theoretical underlying), and not causations.

In addition, we cannot claim that such psychic stress is equal to moral sentiments or psychic costs, but it is not easy to identify alternative reasons; for example, it could be that psychic stress during the decision process is linked to the excitement of evading taxes. However, if this were the case, the predicted outcome would be the opposite, namely a positive correlation between higher excitement

3 Unfortunately, due to the risk of bias, we are unable to administer post-round surveys asking subjects what they were thinking during the decision process. Hence, the underlying thinking process must be treated as a black box.
during the decision period and lower tax compliance. Moreover, it would also be unclear why the maximum psychological response is before and not after the actual decision was made (e.g., waiting for the deterrence process).

![Fig. 1. The Timing of the Decision Process](image)

3. HRV analysis

Much of the literature investigating emotional responses recorded by HRV elicits emotions either by confronting participants with emotionally loaded information or relying on asking subjects to imagine (think of) certain events which are known to be emotionally loaded or to hold emotional connotations for these participants (e.g., Gillie et al. 2015, Rockliff et al. 2008). Indeed, the imagination or thinking of certain events can work even better than inducing emotions by presenting an emotion-eliciting film (see Lane et al. 2009 who used both techniques). In essentially all studies on HRV, including those above, higher HRV is related to negative emotions.

Although previously used in experimental economic research (e.g. Falk et al., 2011; Van Lange et al., 2011; Brandts and Garofalo, 2012), HRV is a measure largely outside the economics toolbox.\(^4\) Rather, HRV is normally used to identify

---

\(^4\) HRV measures have been employed to study responses to social interactions. For example, Falk et al. (2011) evaluated data on 70 subjects, finding that unfair pay (actual share and discrepancy between an actual and appropriate share) is correlated with lower HRV (higher stress) and thus with adverse effects at the physiological level. Van Lange et al. (2011) further find that introducing an experimental protocol measuring heart rate promotes behavioral trust in trust games and reciprocal giving in the trustee, possibly because of interpersonal and intrapersonal mechanisms. On the other hand, Brandts and Garofalo (2012) are unable to show that HRV has a statistically significant effect on performance.
certain medical predispositions (Malik et al., 1996), as well as psychological, emotional, and mental processes (McCraty et al., 1995; Crone, Somsen et al., 2004; Crone, Bunge et al., 2005; Appelhans and Luecken, 2006; Koelsch et al., 2007; Yang et al., 2007).

HRV reflects information about two major parts of the autonomous nervous system (ANS): the sympathetic system, which is responsible for fight-or-flight responses, and the parasympathetic system, which is responsible for rest and relaxation (Appelhans and Luecken, 2006; Levy and Martin, 1979). Changes in the heart rate generated by these two systems occur at different speeds (Appelhans and Luecken, 2006); those induced by the sympathetic system are of considerably longer duration (maximum effect after more than 5 seconds) than those induced by the parasympathetic system (maximum effect after less than 5 seconds) (Levy et al., 1970). This allows decomposition of the HRV signal into a combination of different frequencies with varying relative importance (or spectral power) over time. The frequencies mirror the activation of the sympathetic and parasympathetic systems. Activity in the sympathetic system is mainly reflected by high spectral power in the low frequency band (LF [0.033 - 0.15 Hz]), whereas activity in the parasympathetic system is reflected by high spectral power in the high frequency band (HF [0.15 - 0.4 Hz]) (Malik, 2008).

Although the LF band can be contaminated by parasympathetic activity, the ratio of spectral power in the LF band to that in the HF band can be used as an index of sympathovagal balance (Appelhans and Luecken, 2006), which serves as a useful indicator of psychic stress. Because individuals can react to mental stress with

After exploring how people react to an audience, however, they do observe that, compared with female participants, males are strongly influenced by audience gender. In addition, men exhibit a significant change in blood pressure rather than HRV in reaction to gender pairing in the decision task.
increased sympathetic or decreased parasympathetic activity or both, this LF/HF ratio is a better measure than using the LF or HF alone (Berntson et al., 1994). Its usefulness as an indicator of psychic stress is further supported by evidence that, in controlled experimental environments, the sympathetic-to-parasympathetic ratio generates indicators of the psychological state (Berntson and Cacioppo, 2008; Seong et al., 2004) that correlate with activation in the anterior insular, dorsolateral prefrontal, and anterior cingulate cortices during mental activity. Further detail on HRV, its measurement and data processing is included in the online appendix.

4. Experimental setting

4.1 Experimental procedure

The experiment was conducted over a period of three weeks in April 2009 at the Queensland University of Technology, Brisbane, Australia. Participants were volunteers recruited primarily from first-year economics units using a faculty-wide invitation email and an in-lecture advertisement. The experiment was reviewed by the Queensland University of Technology Faculty Research Ethics Advisory Board which confirmed that it met the requirements of the National Statement on Ethical Conduct in Human Research. The computer-based component was programmed using z-Tree (Fischbacher, 2007). No participant had previous experience with such an experiment. Instructions were displayed on screen\(^5\) and communication with other participants was forbidden during the experiment. Before the experiment began, participants were welcomed and received assistance attaching the heart rate monitors via three single-use electrodes to the chest, connecting the device to the monitor (see online Appendix). Participants were then seated at individual computers.

\(^5\) See online Appendix for the introductory screen.
The experiment, which lasted approximately 90 minutes, proceeded through a number of stages of which the tax compliance stage was the central element. During the tax compliance stage, instructions were also read aloud by a native English speaker, and three test rounds were carried out before the experiment began in order to familiarize participants with the experimental setup. Overall, 14 sessions were conducted covering 45 groups of 4 people (a total of 180 participants).

The first stage of the experiment included a 50 question cognitive skills task on which participants were asked to score as many points as possible during the 12 minutes allocated. This task included logical questions similar to an IQ test. More detail and example questions are included in the online appendix.

The next stage included a risk elicitation task using a multiple price list design as popularized by Holt and Laury (2002), which measures risk as the sum of the number of safe (left-hand side) choices made (see also online Appendix). This method of eliciting risk attitudes has been employed to control for risk attitudes in several previous tax compliance experiments (see, e.g., Heinemann and Kocher, 2013; Cummings et al., 2009). If participants exhibited inconsistencies in the degree of risk aversion (i.e., more than one switching point), we excluded their data from the respective estimations. Accordingly, 36 participants were excluded.

The tax compliance framework was designed to replicate a voluntary income tax reporting decision (see, e.g., Alm, McClelland et al., 1992; Alm, Jackson et al., 1992, 1993). Participants received income over 16 rounds, and then had to declare their income to an experimental “tax authority” and pay taxes on the declared sum (see Figure A1). The after-tax amount went into wealth generation. No tax was paid on undeclared income; however, participants faced the possibility of an audit that would detect and penalize evasion. The tax rate was 30%, and the audit probability was 12.5%. If evasion was detected, evading participants paid the tax on undeclared
income plus a penalty equal this amount (fine rate $f = 2$). These parameters were chosen to mirror the relation between tax compliance and the threat of penalties in various nations. In most countries, the percentage of individual income tax declarations scrutinized in a thorough tax audit is quite small (usually much smaller than the 12.5% chosen here) and the penalty for fraudulent evasion seldom exceeds the amount of unpaid taxes (Alm, et al., 2010).

The experiment implemented two sets of eight rounds of tax declaration decisions (for a total of 16 income declaration decisions). Participants then filled out a post-experimental questionnaire that collected information on demographics, tax morale (see later discussion), and health aspects. From this questionnaire, we determined that our participants were between 17 and 51 years old (average = 23.28, s.d. 5.02) and 56% were female. At the end of the experiment participants were paid according to their earnings during the experiment, receiving between AUD $10 and AUD $29 (mean = AUD $20) based on their accumulated wealth.  

4.2 Experimental design characteristics

We use different design elements that are common to tax compliance experiments (Torgler, 2002, Alm, 1999). An overview of the different treatments as well as the allocation of individuals to different combinations of treatments is provided in Table A1.

4.2.1 Public good structure

4.2.1.1 Traditional public good design structure

In 40 groups the treatments included a public good. The treatments were structured in line with the literature on voluntary public good provision. Specifically, all taxes

6 Conversion rate: AUD $1 = 320 lab$. This information was provided in the instructions.
paid by individuals in a group were accumulated into a group fund, which was multiplied with a factor \( m \) and then equally redistributed to group members (see “transfer income” on the declaration screen in Fig. A2). The redistribution factor corresponds to a fiscal exchange or efficiency parameter in the tax compliance literature. We used three different factors: 0 (no public good), 1, and 2 (the fund is increased by a multiple of 2 before redistribution). Following an approach typical of the tax compliance or public good game (PGG) literature, the main treatment used an efficiency-increasing factor of \( m = 2 \). Hence, tax payments were multiplied by 2 before redistribution to individual players, implying a social efficiency gain through contributions. The amounts collected via audits, in contrast, were not added to the group fund. In the treatments that did not include a public good, participants received nothing in return for their tax payments.

If moral sentiments do not matter, the optimal declared income can be derived in a simple expected utility framework. We use the following equation to express the expected value to risk-neutral taxpayer \( i \) choosing how much income to report (cf. Alm et al., 1999):

\[
EV_i = Y_i - tD_i + ms(tD_i + \sum_j tD_j) - pf[t(Y_i - D_i)]
\]  

(3)

where \( Y_i \) is the income obtained in the experiment before taxation, \( D_i \) is the amount that the individual declares, and \( \sum_j tD_j \) is the sum of taxes generated from the other three members of the group. The variable \( m \), is the surplus multiplier as described above, \( s \) is the individual’s share of the redistributed amount, \( p \) is the probability of detection, and \( f \) is the fine rate on unpaid taxes. If \( pf + ms > 1 \), then the maximization of the expected value \( EV_i \) by given income \( D_i \) indicates full tax compliance as an
optimal strategy. Based on our parameters, however (i.e., $p = 0.125, f = 2, m = 0, 1, 2, \text{and } s = 1/4$), this condition is not satisfied, indicating that zero income reporting would be optimal.

The major advantage of applying this experimental structure in the context of psychic stress or guilt is the ability to explore the consequences of the transfer payment (the amount redistributed). The size of the transfer payment is influenced by the fiscal exchange parameter and by the level of cooperation within the group. If psychic stress is independent of social influence (the behavior of other players in the experiment and the social efficiency of tax payments), any interaction effect between psychic stress and the size of the redistributed amount will be insignificant. On the other hand, a significant interaction effect will indicate that horizontal (through other players and their contributions) and vertical (through the institutional framework and fiscal exchange) social interactions influence psychic stress. We analyze this aspect in the empirical sections.

4.2.1.2 Variation on the traditional public good design

In the framework described above, each individual’s own contribution influences the amount redistributed, thus we also introduced a variation in which the group tax fund was based on the contribution of the other three group members, which reflects the corner case for a very large proportion of taxpayers. Hence, participants’ own contributions did not influence the redistributed sum. This changes equation (3) to equation (4) with $s=1/3$.

$$EV_i = Y_i - tD_i + ms(\sum_{j \neq i} tD_j) - pf[t(Y_i - D_i)]$$ (4)

Note that also in this framework (if $p<0.5$, which is satisfied here) not declaring taxes is optimal for a risk neutral taxpayer.
Participants went through 8 consecutive rounds of both redistribution frameworks described in 4.2.1.1 and 4.2.1.2, comprising a total of 16 periods of tax compliance decisions. (The rule used for the first 8 rounds was determined randomly, except when there was no redistribution at all.) The compliance rate over the 16 rounds between the different redistribution factors is graphed in the online Appendix.

### 4.2.2 Exogenous or endogenous income distribution

Two further experimental treatments used exogenous versus endogenous income distribution. Participants received income at the beginning of each round. In the experiments using endogenous income distribution, the income distribution was based on individual performance in the cognitive skill test. Within every group of four participants, the two with the highest scores in the cognitive skill task received a high income of 400 lab$ each, while the two with the lowest scores received an income of 200 lab$ each (participants were informed about this allocation at the start of the ability test). This income was given in each round but remained the same throughout all 16 rounds. In exogenous sessions, the allocation to high and low income, which also remained the same for all rounds, was made randomly.

### 4.3 Heart rate variability measurements

During the entire experiment, participants’ ECGs were recorded using a Holter Digital ECG recorder (see online Appendix), and from these measurements, an average LF/HF ratio was calculated for each income declaration period. During the recruitment process, participants were informed that the experiment would include the measurement of their heart rate and that they would not be allowed to eat or drink...
anything except water 90 minutes prior to the experimental session. All sessions were held in the afternoon to minimize the effect of daytime variation in heart rate.

4.4 Post-experimental survey

The aim of the post-experimental survey was to collect demographic variables and to measure self-reported tax morale so it could be interacted with psychic stress. To this end, we include two items whose responses can be used as tax morale proxies. The first, and most commonly used (see, e.g., Alm and Torgler, 2006; Torgler, 2007), asks whether cheating on taxes given the chance can always be justified, never be justified, or something in between (measured on a 10-point scale with two extreme points: 1 = “always justified,” 10 = “never justified”). The second, which is closely related to the experimental setting, measures agreement with the statement that taxes are something “taken away” from the individual (1 = strongly agree, 5 = strongly disagree). According to previous research, tax morale is influenced by contextual or external factors such as institutions or culture (Filippin et al. 2013, Torgler 2007). Hence, to improve our analysis of how the role of psychic stress varies under different conditions, we used dummy variables for the different tax morale scales in the expectation that tax morale may measure something other than psychic stress.

5. Empirical results

5.1. Major findings

This section reports the results of analyzing the pooled data from the 16 experimental periods. Our results indicate that tax compliance over time is significant, with a relatively stable rate throughout the experiment (see online Appendix). This stability
implies that our experiment is unaffected by the commonly observed phenomenon that repetition in public good experiments results in decreased cooperation, a finding that has intensified discussion on whether confusion and inexperience are driving results (Andreoni, 1995). Whereas average compliance is between 50% and 60% throughout the experiment, there is also a large degree of heterogeneity, including significant proportions of data at the corners of the decision space (i.e., full evasion and full compliance, see online Appendix). We account for this by using tobit models capable of handling clumps of data at these two endpoints. We also need to account for the panel structure of the data, but, as we are interested in the influence on tax compliance of certain individual-specific characteristics, we cannot use individual-fixed effects. Rather, we employ group-fixed and time effects and cluster standard errors by individual.

To explore the link between psychological strain and tax compliance, we assume the following basic relation:

$$Compliance_{it} = F[(LF/HF)_{it}; X_{it}; TD_t; GD_i]$$

where $Compliance_{it}$ is the proportion of income declared (ratio of declared to actual income). Since timing is crucial in this setting we provide results from two different approaches to physiological data matching. Table 1 contains our regression results. Firstly, we use the average LF/HF ratio from the moment the income declaration screen appears to the moment the participants click the OK button after declaring their income (specifications (1) to (3)). Secondly, we use an optimal time window (specifications (4) and (5)) based on a maximum likelihood estimation detailed further below. We control for several factors ($X_{it}$), including the accumulated wealth at the time of the decision ($Log(wealth)_{it}$); received back-transfers from
redistribution in the previous round \( (\text{Transfer payment}_{(t-1)}) \), which captures a participants’ experimental history; cognitive skills \( (\text{Cognitive skills}) \); whether or not a participant was audited \( (\text{Audited}_{(t-1)}) \) or punished in the previous round \( (\text{Fine}_{(t-1)}) \), and the number of times fined during the experiment \( (\text{Times fined}) \), which takes into account potential longer lasting effects. We first report results without these three deterrence variables as one could argue that a deterrence experience may add to stress, rendering it endogenous (see specification 1).

We also control for other individual characteristics identified in the literature as driving forces in compliance (see Torgler 2006, 2007), including gender (dummy for \( \text{Female} \)), age, risk aversion, and church attendance (\( \text{Church attendance}: 1 = \text{once a week or more than once a week}, 0 \text{ otherwise} \)). Finally, we control for the experimental design and group dynamics (group dummies, \( GD_i \)), and time dynamics (time dummies, \( TD_t \)).

Our estimations indicate a positive and statistically significant relation between tax compliance and psychic stress during the tax declaration process. For specifications (1) and (2), the coefficient is statistically significant at the 5% level, while the marginal effect at the mean demonstrates that a one-unit increase in the LF/HF ratio parallels an increased tax compliance between 1.0 and 1.9 percentage. The LF/HF variable has a standard deviation of 1.68 in (2) which implies a change of roughly 0.6 percentage points in tax compliance per standard deviation change in LF/HF. Specification (3) incorporates the lagged variable LF/HF\(_{(t-1)}\) to account for the fact that psychic stress from the former period could influence the current decision. The lagged LF/HF value is also statistically significant but only at the 10% level. Similarly, in specification (5) the lagged LF/HF value is statistically significant at the 10% level, but its effect is substantially smaller than the LF/HF proxy during the actual decision process.
In specifications (4) and (5), we implement an alternative LF/HF proxy using a different approach to choose the most appropriate time interval. This estimation requires a two-step procedure. The first step is a maximum likelihood estimation of a reduced form tobit model using a data set containing one-second interval LF/HF indicators from the decision period until it was submitted. This approach ensures that, when matching the LF/HF ratio to the decision, the model is able to select the most appropriate time offset for each individual (see online Appendix for further clarification). In the second step we then use the parameters from the first regression to construct an adjusted LF/HF proxy, which is used in the second step tobit estimations (see specifications (4) and (5)). The average estimated time offset is -16.81 seconds, which, as mentioned previously, indicates that (on average) participants experienced the strongest impact of psychic stress approximately 17 seconds before submitting their income declaration. Therefore, this procedure provides the best time interval for the maximum correlation of physiological stress with tax compliance, with a marginal effect of 12.6 percentage points as shown in specification (5). As this proxy has a lower standard deviation (0.66), this equates to a 19.1 percentage point increase in tax compliance related to a one standard deviation increase in this LF/HF proxy.\textsuperscript{8} Taken together, the above results provide clear evidence that psychic stress is significantly related to tax compliance. Specifically, the decision to pay more taxes (and cheat less) is correlated with psychic stress during the tax declaration decision.\textsuperscript{9}.

\textsuperscript{8} It should be noted that if there were no sustained effect, the maximum likelihood would render the parameter out of bounds for all observations. In other words, if there were no effect, the adjusted LF/HF proxy would be zero for all observations (i.e., no “real” effect). One desirable side effect of this procedure is that it allows the offset parameter to lie outside the 120-second interval, which, because there are no available data, effectively reduces the influence of the LF/HF parameter to zero for those individuals. As a result, no LF/HF ratio effect is implied for about 60% of the sample.

\textsuperscript{9} 324 observations showed HRV data points within the decision time interval that were abnormal. These observations were excluded from the analysis for specifications (1) to (3) to reduce noise in the estimates. To reduce the impact of outliers, we first excluded faulty recordings by manual inspection.
Table 1
Relation between psychic stress and tax compliance.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable: tax compliance rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Independent variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychic stress</td>
<td>0.025**</td>
<td>0.021**</td>
<td>0.018*</td>
<td>0.282***</td>
<td>0.227***</td>
</tr>
<tr>
<td>(LF/HF),t</td>
<td>(2.24)</td>
<td>(2.00)</td>
<td>(1.87)</td>
<td>(5.00)</td>
<td>(4.99)</td>
</tr>
<tr>
<td></td>
<td>0.019</td>
<td>0.010</td>
<td>0.008</td>
<td>0.126</td>
<td>0.102</td>
</tr>
<tr>
<td>Psychic stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(LF/HF),t1</td>
<td>0.019*</td>
<td>0.127***</td>
<td>0.126***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deterrence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audited,(t-1)</td>
<td>0.032</td>
<td>0.012</td>
<td>0.077</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.37)</td>
<td>(0.14)</td>
<td>(0.79)</td>
<td>(0.68)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.015</td>
<td>0.006</td>
<td>0.034</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Fine,(t-1)</td>
<td>-0.002***</td>
<td>-0.002**</td>
<td>-0.003***</td>
<td>-0.003***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.94)</td>
<td>(-2.53)</td>
<td>(-2.76)</td>
<td>(-2.76)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Times fined</td>
<td>-0.084**</td>
<td>-0.085**</td>
<td>-0.137***</td>
<td>-0.135**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.08)</td>
<td>(-2.10)</td>
<td>(-3.08)</td>
<td>(-3.08)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.039</td>
<td>-0.039</td>
<td>-0.061</td>
<td>-0.060</td>
<td></td>
</tr>
<tr>
<td>Public good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer payment,(t-1)</td>
<td>-0.003</td>
<td>-0.003***</td>
<td>-0.003***</td>
<td>-0.002***</td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td>(-3.11)</td>
<td>(-3.01)</td>
<td>(-3.06)</td>
<td>(-2.23)</td>
<td>(-2.08)</td>
</tr>
<tr>
<td></td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Wealth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(wealth),t1</td>
<td>-0.630***</td>
<td>-0.687***</td>
<td>-0.682***</td>
<td>-0.991***</td>
<td>-1.027***</td>
</tr>
<tr>
<td></td>
<td>(-2.79)</td>
<td>(-3.11)</td>
<td>(-3.15)</td>
<td>(-4.89)</td>
<td>(-5.13)</td>
</tr>
<tr>
<td></td>
<td>-0.480</td>
<td>-0.317</td>
<td>-0.316</td>
<td>-0.443</td>
<td>-0.459</td>
</tr>
<tr>
<td>Individual characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive skills</td>
<td>0.047***</td>
<td>0.048***</td>
<td>0.047***</td>
<td>0.050***</td>
<td>0.051***</td>
</tr>
<tr>
<td></td>
<td>(4.25)</td>
<td>(4.47)</td>
<td>(4.43)</td>
<td>(5.16)</td>
<td>(5.34)</td>
</tr>
<tr>
<td></td>
<td>0.036</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
<td>0.023</td>
</tr>
<tr>
<td>Age</td>
<td>0.036***</td>
<td>0.035***</td>
<td>0.036***</td>
<td>0.046***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.58)</td>
<td>(3.57)</td>
<td>(3.70)</td>
<td>(4.81)</td>
<td>(4.84)</td>
</tr>
<tr>
<td></td>
<td>0.028</td>
<td>0.016</td>
<td>0.017</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>Female</td>
<td>0.365***</td>
<td>0.330***</td>
<td>0.330***</td>
<td>0.188***</td>
<td>0.190**</td>
</tr>
<tr>
<td></td>
<td>(3.44)</td>
<td>(3.28)</td>
<td>(3.37)</td>
<td>(2.21)</td>
<td>(2.25)</td>
</tr>
<tr>
<td></td>
<td>0.278</td>
<td>0.152</td>
<td>0.153</td>
<td>0.084</td>
<td>0.085</td>
</tr>
<tr>
<td>Church attendance</td>
<td>2.104***</td>
<td>1.960***</td>
<td>1.910***</td>
<td>1.441***</td>
<td>1.449***</td>
</tr>
<tr>
<td></td>
<td>(5.91)</td>
<td>(6.32)</td>
<td>(6.20)</td>
<td>(4.30)</td>
<td>(4.19)</td>
</tr>
<tr>
<td></td>
<td>1.604</td>
<td>0.903</td>
<td>0.885</td>
<td>0.645</td>
<td>0.648</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>0.028</td>
<td>0.034</td>
<td>0.031</td>
<td>0.047**</td>
<td>0.048**</td>
</tr>
</tbody>
</table>

and then removed the top and bottom 5% percentiles from this data over the whole sample. However, the results remain robust with those 324 observations included. Specifications (4) and (5) were not affected by it as they covered a more precise and smaller time interval and the irregularities did not affect the results in any way.
With respect to the control variables, women are more compliant than men, and age is positively correlated with compliance, although this latter finding should be treated with caution because most of our participants were relatively young. Higher cognitive skills, a higher level of religiosity and more risk aversion are also associated with higher tax compliance, although the coefficient for risk aversion is only statistically significant at the 10% level in specifications (3) and (4). Neither prior audit, size of the penalty paid, nor number of past fines seems to improve tax compliance in the subsequent period. On the contrary, compliance is negatively related to both fine proxies. These results, however, cannot be compared to the available evidence in the tax compliance literature because the latter focuses primarily on fine rate changes and differences in audit probability (Alm, 1999), two variables that remained constant in our experimental design. Nevertheless, these outcomes are comparable to the results of experimental studies that use similar proxies (Torgler and Schaltegger, 2005).

We also identify a negative relation between tax compliance and accumulated wealth, and between tax compliance and redistributed amounts (transfer payments) in the previous round, which indicates that compliance within groups is not reciprocal. Interestingly, however, this effect is only statistically significant for the
redistribution case in which the individual contribution has no impact. When the individual contribution matters, the coefficient is no longer statistically significant ($t = -0.22$). We also explore whether psychic stress is influenced by reciprocal considerations. Two potential hypotheses related to such a link are that (i) the higher the group’s past compliance level (e.g., in the former period), the higher the moral costs of noncompliance and (ii) the higher the fiscal exchange parameter, the more money the subjects receive because of multiplier effects. However, our data indicates (results not reported here) no change in the influence of psychic stress because of these factors, suggesting that psychic stress is internal and not driven by external factors. Similarly, the relationship between psychic stress and tax compliance appears to be unaffected by whether the income distribution is exogenous or endogenous.

Table 2
Tax morale and psychic stress.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Dependent variable: tax compliance rate</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coeff.</td>
<td>t-stat.</td>
</tr>
<tr>
<td>Psychic stress (LF/HF)</td>
<td>-0.038</td>
<td>(-1.60)</td>
<td>-0.087*</td>
</tr>
<tr>
<td>Tax morale value 2</td>
<td>0.181</td>
<td>(0.43)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 3</td>
<td>0.038</td>
<td>(0.13)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 4</td>
<td>0.344</td>
<td>(1.18)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 5</td>
<td>0.251</td>
<td>(1.00)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 6</td>
<td>0.582***</td>
<td>(2.26)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 7</td>
<td>-0.014</td>
<td>(-0.05)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 8</td>
<td>0.385*</td>
<td>(1.74)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 9</td>
<td>0.242</td>
<td>(0.97)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 10</td>
<td>0.542*</td>
<td>(1.87)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 2 * Psychic stress (LF/HF)</td>
<td>0.170</td>
<td>(0.93)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 3 * Psychic stress (LF/HF)</td>
<td>0.170***</td>
<td>(3.06)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 4 * Psychic stress (LF/HF)</td>
<td>0.025</td>
<td>(0.69)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 5 * Psychic stress (LF/HF)</td>
<td>0.096**</td>
<td>(2.33)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 6 * Psychic stress (LF/HF)</td>
<td>0.046</td>
<td>(1.31)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 7 * Psychic stress (LF/HF)</td>
<td>0.024</td>
<td>(0.52)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 8 * Psychic stress (LF/HF)</td>
<td>0.068**</td>
<td>(2.24)</td>
<td></td>
</tr>
<tr>
<td>Tax morale value 9 * Psychic stress (LF/HF)</td>
<td>0.079**</td>
<td>(2.36)</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Coefficient</td>
<td>Standard Error</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>1st day tax morale value 10 * Psychic stress (LF/HF)</td>
<td>0.069**</td>
<td>(2.33)</td>
<td></td>
</tr>
<tr>
<td>2nd tax morale proxy value 2</td>
<td>-0.507***</td>
<td>(-2.83)</td>
<td></td>
</tr>
<tr>
<td>2nd tax morale proxy value 3</td>
<td>-0.550**</td>
<td>(-2.51)</td>
<td></td>
</tr>
<tr>
<td>2nd tax morale proxy value 4</td>
<td>-0.379**</td>
<td>(-2.03)</td>
<td></td>
</tr>
<tr>
<td>2nd tax morale proxy value 5</td>
<td>-0.230</td>
<td>(-0.88)</td>
<td></td>
</tr>
<tr>
<td>2nd Tax morale value 2 * Psychic stress (LF/HF)</td>
<td>0.109**</td>
<td>(2.07)</td>
<td></td>
</tr>
<tr>
<td>2nd Tax morale value 3 * Psychic stress (LF/HF)</td>
<td>0.094</td>
<td>(1.64)</td>
<td></td>
</tr>
<tr>
<td>2nd Tax morale value 4 * Psychic stress (LF/HF)</td>
<td>0.146***</td>
<td>(2.74)</td>
<td></td>
</tr>
<tr>
<td>2nd Tax morale value 5 * Psychic stress (LF/HF)</td>
<td>0.136**</td>
<td>(2.50)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Tobit model with limits at 0 and 1 for the dependent variable: the tax compliance rate as the ratio of reported to true income. Standard errors are clustered by individual. Significance levels: *0.05 < p < 0.10, **0.01 < p < 0.05, ***p < 0.01.

5.2 Tax Morale and psychic stress

Next, we explore whether the effect of psychic stress may differ depending on a subject’s tax morale using the two tax morale variables introduced in the last section.

There could be a non-linear relationship between tax morale and psychic cost as tax morale and psychic stress could vary for different types of individuals. The results, reported in specifications (6) and (7) of Table 2, indicate that psychic stress has a consistently strong and positive influence on tax compliance for subjects with high tax morale, an outcome that is robust across the two different tax morale proxies. We also note that psychic stress increases compliance for subjects whose tax morale values lie in the middle or lower middle range (3 and 5 for proxy 1; 2 for proxy 2).

Hence, overall, these results indicate that individuals with different levels of tax morale respond differently with respect to psychic stress and the corresponding tax compliance. More specifically, these results imply that tax compliance due to tax
morale may indeed be channeled through the imposition of psychic costs, particularly for those who think that tax evasion is morally questionable.\textsuperscript{10}

5.3 Different types of taxpayers

The interaction effects observed previously and the large number of censored observations (particularly when using our alternative LF/HF proxy) indicates that there are different types of taxpayers with respect to tax compliance, tax morale, and their corresponding emotional reactions. To explore further whether taxpayers differ, we apply a $K$-medians partition\textsuperscript{11} cluster analysis with absolute-value distance $\sum_{k=1}^{P}|x_{ki} - x_{kj}|$\textsuperscript{12} as a similarity/dissimilarity measure for continuous data using tax compliance, the two tax morale proxies, and the main HRV variable. As indicated in Table 3, there seem to be three types of taxpayers: (i) those with lower tax morale and tax compliance and a level of psychic stress that ranges between those of the two other groups (cluster 1), (ii) those with high tax morale and tax compliance but high psychic stress (cluster 2), and (iii) those with high tax morale and tax compliance but no psychic stress (cluster 3). Interestingly, both clusters of taxpayer with high tax morale have almost exactly the same rate of tax compliance, regardless of whether we look at the mean or the median.

Table 3

Types of taxpayers.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Percent</th>
<th>Tax compliance</th>
<th>LF/HF ratio</th>
<th>Tax morale proxy 1</th>
<th>Tax morale proxy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>39.33</td>
<td>0.416</td>
<td>1.622</td>
<td>4.624</td>
<td>2.690</td>
</tr>
</tbody>
</table>

\textsuperscript{10} While the results seem to be robust with respect to the measure of tax morale used, we agree with the point made by one of the referees that the variation in the estimates indicate also the level of noise involved in this analysis.

\textsuperscript{11} The $k$ clusters are formed by an iterative process; the medians are computed to represent the group centres at each step (i.e., observations are assigned to groups based on closeness).

\textsuperscript{12} This examines the absolute differences between coordinates of a pair. $i$ and $j$ define the different subjects while $k$ the features or variables ($p$=number of variables).
26.09 0.613  3.065  8.708  3.132
37.58 0.613  3.065  8.193  3.027
                  Median
   1  39.33  0.30  1.160  5    2
   2  23.09  0.75  2.453  9    3
   3  37.58  0.75  0.603  8    3

5.4 Robustness checks

As a first robustness check, we conduct separate analyses of the periods with
different public good redistribution rules. Applying specification (1) indicates that in
both cases the coefficient \((\text{LF/HF})_t\) remains statistically significant. We then check
whether there is a possible nonlinear relation between psychic costs and tax
compliance. After adding \((\text{LF/HF})_t^2\) into the two specifications reported in Table 1,
we observe that the coefficient \((\text{LF/HF})_t^2\) is negative while \((\text{LF/HF})_t\) remains positive
with a \(t\)-value of 2.20 (indicating that tax compliance is increasing at a decreasing
rate with an increase in psychic stress). The coefficient for \((\text{LF/HF})_t^2\), however, is not
statistically significant \((t = -1.53)\), indicating that the linear parameter is a sufficient
fit. As further robustness checks, we add controls for HRV-related health factors
such as alcohol consumption (drinks per week), smoking habits (“daily = 3,” “less
frequently = 2,” “not at all” = 1) and physical/sport activity (minutes per week).
These additions, however, barely change the impact of the LF/HF ratio on tax
compliance \((t = 2.03, \text{ marginal effect} = 0.01)\).\(^{13}\)

5.5 Psychic stress after noncompliance

Finally, we examine whether individuals experience psychic stress after committing
an illegal act (remorse) or after getting caught (shame).\(^{14}\) The feeling of remorse is
related to what Cooter and Ulen (2004, p. 464) call “Saturday night fever”: waking

\(^{13}\) Regression tables are in the online Appendix.
\(^{14}\) For detailed results see online Appendix.
up the next morning saying, “I can’t believe what I did last night!” Our results reveal a negative but insignificant ($t = -0.80$) correlation between tax compliance in a previous period ($t-1$) and psychic stress in the current period ($t$). We therefore refine the analysis to explore the relation between tax compliance and psychic stress measured from the moment the auditing screen appears until the moment the participant clicks OK (see Figures 1 and A2). The results indicate that in this case, tax compliance is also negatively but insignificantly ($t = -1.18$) related to psychic stress.

We also look at the relation between deterrence in the previous period ($t-1$) and psychic stress in the current round ($t$). We find no significant relation between psychic stress and being audited ($t = -1.15$) or the size of the fine ($t = -1.10$), which indicates little psychic stress after the individual experiences punishment. Similarly, when we examine how psychic stress during the audit screen changes with the number of times fined, the correlation is negative (i.e., the more times fined, the less the psychic stress), but the coefficient is again not statistically significant (-0.51).

### 6 Conclusions

Our results show that the analysis of tax compliance may benefit from a deeper understanding of moral sentiments and its physiological foundation. Contrary to the findings of Coricelli et al. (2010) but in line with Erard and Feinstein’s (1994) theoretical framework, we find that higher psychic stress increases tax compliance. We therefore hypothesize that this higher psychic stress may be due to an intention or contemplation of making a false declaration, meaning that the physiological

---

15 Based on OLS estimations with group and time fixed effects and standard errors clustered by individual.
16 For these estimates, we use the same OLS specification (group and time dummies and clustering).
17 We explore deterrence using two separate regressions.
18 The OLS regression is structured as before.
measurement could be an indicator of moral sentiments or psychic costs. However, future studies could try to manipulate moral sentiments such as guilt or manipulate stress to see whether there is a causal relationship.

Based on additional analyses, we can also classify taxpayers into three types: those with high tax morale and tax compliance but no psychic stress, those with high tax morale and high tax compliance but high psychic stress, and those with lower tax morale and tax compliance and a psychic stress level somewhere between those of the other two groups. This suggests an interpretation of our measures of psychic stress: they are being triggered by moral emotions, which motivate compliance decisions. Future research could explore in more detail the implications of different types of taxpayers. There are already attempts in the literature to classify taxpayers into different types (Vogel 1974, Torgler 2003).

Our results may also have implications for research beyond tax compliance, such as contributions to the commons (Dietz et al., 2003), which underlie many public good activities. However, further research is required to determine if our findings on the role of the physiologically measurable psychic costs of moral sentiments will hold independent of a specific domain. Furthermore, our results could have implications for policy; for example by documenting the importance of increasing awareness regarding the value of taxes, particularly with respect to their use. In addition, it may increase awareness of the commons more generally, which could foster compliance. From this perspective, understanding the social and economic conditions under which contribution elicits more or less psychic strain may be a way to improve tax compliance and, if results apply more generally, to public good frameworks and the effectiveness of public good contribution systems in general.
Finally, the present study demonstrates that HRV data can be used as a biomarker to observe decision behavior influenced by the nonmonetary and noneconomic incentives (like norms) underlying contributions to the commons or society in general. Not only can such data help to predict and explain participants’ behavior, but carefully designed experiments that record individuals’ HRV can help to shed light on latent processes by offering objective data where subjective motivations like moral sentiment and social norms affect decision making.

References


Cooter, R., Ulen, T., 2044. Law & Economics. Pearson Addison Wesley, Boston.


Appendix A. Additional tables and figures

**Fig. A1.** Distribution of the LF/HF ratio.

**Table A1.** Treatments and observations.
### Treatments

<table>
<thead>
<tr>
<th>Multiplication factor ((m))</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>endogenous</td>
<td>exogenous</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>20</td>
<td>64 (28, 36)</td>
<td>68 (28, 40)</td>
</tr>
<tr>
<td>Participants with usable HRV data</td>
<td>19</td>
<td>64 (28, 36)</td>
<td>67 (27, 40)</td>
</tr>
</tbody>
</table>

**Notes:** The table shows the tax rate for compliant subjects and the number of participants in each treatment. The numbers in brackets specifies the number of participants which first played the inclusive (included own contributions) or the exclusive treatment (excluding own contributions) (inclusive first, exclusive first). The table also includes the respective number of participants with usable HRV measures. The main treatment, to which other treatments are variations for comparison, is highlighted in bold.

---

**Fig. A2.** Income tax declaration screen.

Notes: Because the time data on events in the computer-based experiment are prone to variation caused by network traffic issues, the time recorded only gives an indication of when the event actually happened (i.e., when the screen with the information was actually visible to the participant). We therefore isolated the laboratory network from the university network and assigned it to its own subnet, which substantially reduced the recording discrepancy (to <6 milliseconds).
Research Highlights: Dulleck, Fooken, Newton, Ristl, Schaffner, Torgler: Tax Compliance and Psychic Costs: Behavioral Experimental Evidence Using a Physiological Marker

This article presents research that is innovative with respect to the following:

• It uses heart rate variability, an established measure of stress, in a taxation experiment.
• It adds to the very limited literature providing physiological evidence that moral sentiment or social norms (tax moral) matter for peoples’ decision to pay taxes.
• It presents experimental data indicating that feeling of guilt may explain compliant behaviour.
• It is first to look at heterogeneity among experimental tax payers with respect to compliance behavior and physiological reactions.