“Willing to Pay?” Tax Compliance in Britain and Italy: an Experimental Analysis

Nan Zhang¹,*, Giulia Andrighetto¹,², Stefania Ottone³, Ferruccio Ponzano⁴, Sven Steinmo⁵,6,*

1 Department of Social and Political Sciences, European University Institute, San Domenico di Fiesole (FI), Italy
2 Institute of Cognitive Science and Technology, National Research Council, Rome, Italy
3 Department of Economics, Management and Statistics, University of Milano-Bicocca, Milano, Italy
4 Department of Law, Politics, Economics and Social Sciences, University of Eastern Piedmont, Alessandria, Italy
5 Robert Schumann Center for Advanced Studies, European University Institute, San Domenico di Fiesole (FI), Italy
6 Department of Political Science, University of Colorado, Boulder, USA

* E-mail: nan.zhang@eui.eu
* E-mail: sven.steinmo@eui.eu

Abstract

As shown by the recent crisis, tax evasion poses a significant problem for countries such as Greece, Spain and Italy. While these societies certainly possess weaker fiscal institutions as compared to other EU members, might broader cultural differences between northern and southern Europe also help to explain citizens’ (un)willingness to pay their taxes? To address this question, we conduct laboratory experiments in the UK and Italy, two countries which straddle this North-South divide. Our design allows us to examine citizens’ willingness to contribute to public goods via taxes while holding institutions constant. We report a surprising result: when faced with identical tax institutions, redistribution rules and audit probabilities, Italian participants are significantly more likely to comply than Britons. Overall, our findings cast doubt upon “culturalist” arguments that would attribute cross-country differences in tax compliance to the lack of morality amongst southern European taxpayers.

Introduction

Modern welfare states face a set of difficult challenges as they adapt to the demographic, economic and political strains of the early 21st century. States must struggle to maintain adequate support for social welfare and educational programs in the face of growing distrust of bureaucratic institutions, intense pressures to cut taxes for politically powerful constituencies, and fiscal burdens arising from an aging population. The ability of governments to collect revenues in an efficient and cost-effective manner is of central importance to how successfully states meet their policy goals. And to ensure a healthy fiscal foundation, states must be able to control (or reduce) tax evasion on the part of their citizens.
Yet, while Western European states generally possess tax systems sharing many of the same formal features, actual rates of tax compliance vary widely across these societies. Moreover, evasion rates also seem to follow a geographic pattern, with high levels of compliance in northern Europe, and widespread under-reporting in the countries further south. Using the size of the “shadow economy” as a proxy for tax evasion, Schneider and Enste find the lowest compliance rates in Western Europe in Portugal, Spain, Italy and Greece.

The literature has advanced several “institutionalist” theories to account for this cross-national variation in tax compliance. In large part, these explanations focus on the relationship between the quality of government and citizens’ willingness to comply with fiscal demands. Specifically, this literature argues that citizens are more likely to pay their taxes if they believe that the government is spending their money honestly and efficiently. By contrast, when citizens perceive public institutions as corrupt and wasteful, they are likely to reciprocate by being dishonest in turn. Thus, one explanation for EU-wide differences in levels of tax compliance is that, in southern European countries, people often interact with low quality institutions for which they are (unsurprisingly) unwilling to pay.

A second set of theories links tax compliance to broader cultural norms and values. Within Europe, one important axis of cultural variation concerns how different societies draw the boundaries of moral behavior. Specifically, southern European societies are typically characterized as more “familistic” or “collectivist,” and ethical conduct is often assumed to apply only to a small circle of familial or personal relationships, while outside of this circumscribed network, selfish or opportunistic behavior is norm. In the public sphere, individuals follow the rules not out of some internalized sense of “right” and “wrong,” but only when they are coerced to do so. By contrast, northern European societies are often said to emphasize values of “autonomy” or “individualism,” and citizens are presumed to apply the same ethical principles that prevail within familial relations to conduct in the civic realm.

This distinction between what Tabellini has termed “limited” and “generalized” morality has direct implications for the level of tax compliance within a society. As many scholars have noted, in most cases, audit rates and punishment probabilities are insufficient to deter cheating, and states must rely upon voluntary compliance to collect fiscal dues. However, if “cheating the system” imposes little moral cost, then the willingness to pay is undermined. Thus, a broader cultural argument would lead us to expect greater tax evasion in southern European countries, independent of institutional performance.

Our research attempts to test this hypothesis using cross-cultural behavioral experiments. The advantage of this approach is that it allows us to hold formal institutions (e.g. tax rates, audit probabilities, the efficiency of the state, etc.) constant across countries, and thereby isolate the influence of broader cultural factors on fiscal behavior. In this paper, we report results from two countries - the UK and Italy - which we take as “representatives” of northern and southern Europe. For example, Slemrod estimates the evasion rate in the UK to be around 8% or 9% of GDP, while comparative figures for Italy can reach as high as 25% to 30%. Culturally, Italy is often vilified - both in the press as well as in popular opinion - as the quintessential “amoral” society in which people cannot be trusted to behave ethically outside the

---

1The literature here is vast, and concerns not only the quality of institutions. For example, other scholars have noted the importance of factors such as participation in direct democracy or “fair treatment” at the hands of the tax authorities. However, we have no evidence that these factors vary systematically between northern and southern Europe.

2Putnam has also argued for the importance of social capital for democratic governance more generally.

3Several previous papers have employed this strategy in the tax compliance context.
network of familial and personal relations. By contrast, Britain is rather typical of Protestant, northern European societies in terms of cross-national rankings of “autonomy” and “individualism”.

Our tax experiment involves over 500 participants across multiple locations in Italy and the UK. The main experimental task consists of a tax compliance scenario in which participants earning real money are asked to report their income under a variety of tax rates and redistributive rules. By comparing income declarations across countries, we are able to investigate whether, independently of the institutions, Italians are indeed less compliant than Britons when faced with identical fiscal choices.

To preview our results, we find no evidence to support the contention that the morality of tax compliance is weaker in Italy, compared to the UK. Instead, we find that the compliance rate amongst Italian participants is significantly higher than amongst Britons. These results remain robust to the inclusion of a host of demographic controls, and are reproduced in multiple experimental locations in the two countries. In summary, although stereotypes about the “amorality” of Italian (and, more generally, southern European) taxpayers abound in the popular consciousness, our results suggest that cultural values cannot explain the significant cross-national variation in evasion rates that we observe in the real world. In the concluding section, we discuss several implications of our findings for future work.

Methods

Overview

Our experiments were conducted at six universities across the United Kingdom and Italy at various points during the academic year 2013-2014. Our team spent over a year designing and re-designing our experimental protocols to ensure the consistency of the laboratory set-ups and selection pools in each of these locations. Everything from the recruitment methods, to the way participants entered the lab, to the final payments procedure was the same in each session. Our experimental instructions (both oral and on-screen) were also translated and back-translated between Italian and English by different native speakers to ensure consistency in meaning.

Each of the universities in which we conducted our study maintains an electronic database of individuals who had expressed interest in participating in behavioral experiments. These participant pools are composed mainly of undergraduate students, but also include a number of non-students and people who had already graduated. Several days prior to the actual sessions, individuals in the database receive an email informing them of the opportunity to take part (for a reasonable hourly wage) in an upcoming research project. The email also contained a link where individuals could sign up for one session of their choice from a list of scheduled sessions.

On the day of the experiment, upon arrival in the laboratory, participants are given a randomly-drawn, anonymized ID number and assigned to a corresponding personal computer terminal. Participants undertake all experimental tasks via computer, and the terminals are partitioned to ensure that participants could not communicate during the session, nor observe what others are doing. Also, to ensure anonymity, we announced that decisions and payments would be linked only to participants’ ID-numbers, and not to individual names.

To respect the anonymity of participants, we do not report the experimental locations in this paper. Details are available from the authors upon request. The experiments are programmed in zTree.

Participants were all recruited using ORSEE.

Only the participants and the researchers are present for each session.
Once all participants have been seated at their individual terminals, we begin the session by reading a short introductory script. Participants are informed that they would be asked to complete a number of tasks (which we would gradually describe to them) and make a number of choices. Based on their choices and the choices of the other participants, they would earn experimental currency units (ECUs), which would be converted into real money at the end of the session. Individuals then take part in the tax compliance experiment, which we describe in the following section. Finally, participants complete an anonymous computerized survey designed to collect demographic and attitudinal information before receiving payment anonymously in cash. In all, each session of the experiment lasts about 90 minutes, and subjects earn an average of 16 euros / 14.5 pounds for their time.

Importantly, prior to beginning the experimental tasks, participants are given no information about the aims of the research project, nor about the types of information and behavior we seek to elicit. Also, at no point whatsoever are participants told that they are taking part in a larger study comparing decision-making across national groups. Finally, when moderating each session, we made sure to employ only native speakers (and, in the Italian case, speakers with the “correct” regional accent). These procedures are implemented to ensure that participants would not be subject to national (or group-level) reputational concerns when making their decisions.

Ethics Statement

Our experiments have been approved by the IRB Committee at the University of Colorado, Boulder, where the principal investigator holds a professorship. Our project has also been approved by the Ethics Council of the European Research Council, and the European University Institute Ethics committee. Finally, our work has also been authorized by all of the Italian and UK laboratories we have used, but we did not undergo a separate university-based IRB review in these cases. All participants signed a written consent form prior to taking part in the study.

Participants

The data we present in this paper are drawn from 31 different experimental sessions involving a total of 671 participants from a variety of academic disciplines. Because we are interested in comparing specifically British and Italian participants, we retain the data for only native students, whom we define as those individuals born in Britain (Italy) to British (Italian) parents. The result leaves us with a subset of 531 participants, of which 281 (52.9%) are from Italy and 250 (47.1%) are from the UK.

7 These units are converted into local currencies (pounds and euros) so that at the end of the experiment, the average participant would receive an income of approximately twice the average hourly wage for student employment in the local context.

8 Answers to the questionnaire are linked to decisions in the experiment via the anonymized ID numbers. We do not collect information in the questionnaire which would allow us to identify individual participants by name. A text version of the computerized questionnaire is included in the supplementary materials.

9 For example, if Italians had known that their behavior would be compared to the behavior of British participants, they might have made efforts to look more honest, in order to improve the reputation of Italy in the world. We are grateful for an anonymous reviewer for highlighting this point.

10 Participants who wish to leave the experiment early could elect to receive a 5 euro / pound show-up fee. In practice, no individuals refused to participate or dropped out.

11 Because of a misunderstanding at the recruitment stage, one session in [location anonymized] enrolled many immigrants. While we excluded the immigrants’ data, we were also concerned that interacting with a disproportionate number of foreigners may have skewed the behavior of even native-born participants. We therefore drop this session entirely from the analysis.

12 Including the entire sample of 671 participants does not substantively change our results.
**SI Table** presents descriptive statistics for our participant pools. Overall, 56% of our participants are male, with an average age of 23.8 years (s.d. = 7.7 years).

We note that there are several slight demographic differences between our British and Italian samples. In particular, British participants are significantly more likely to be employed and to report a higher willingness to take risks, while Italians are more likely to study economics. In addition, British participants tend to be more skilled at our real effort task (described below), and hence earn more ECUs in the experiment. On the other hand, we detect no significant differences between the two populations in terms of gender, age, or previous participation in experiments. We control for all of these demographic characteristics in our analyses below.

**Tax Compliance Experiment**

**Design and Procedure**

The tax compliance experiment proceeds as follows. Participants begin by completing a clerical task, in which they must copy rows of information from a sheet of paper into the computer. For each correctly copied row, participants earn 10 ECUs. Next, participants are asked to declare this income for taxation purposes under different scenarios (an example of the income reporting screen is shown in **S1 Fig**). In the terminology of the experiment, each scenario constitutes a “round.” Participants are informed that they are free to declare any amount of their income - from 0% to 100% - in each round, with the knowledge that they would pay taxes on only the reported portion of their incomes.

In all, the experiment is composed of nine separate income declaration rounds [13]. In each round, we specify different rules for the taxation and redistribution of declared incomes (which we describe shortly). In doing so, we wanted to explore the effects of different causal variables from the tax policy literature. As a consequence, readers will note that our design is similar to previous tax compliance experiments that have also investigated the influence of these variables [32, 42–48]. Importantly, our primary contribution examines how, holding these institutional features constant, tax compliance varies across countries [14, 36].

In Rounds 1 through 3, we implement a flat 30% tax rate, and vary how tax revenues are redistributed to participants, thus simulating behavior under different levels of “efficiency” in providing public goods. In Round 1, there is no redistribution, and all tax revenues are “wasted.” In Round 2, all taxes collected are redistributed on an equal basis to all participants. Finally, in Round 3, all taxes collected are first doubled and then redistributed equally to all participants.

By contrast, in Rounds 4 through 6, we hold redistribution constant and instead vary the tax rate. We use a flat tax rate of 10% in Round 4, of 30% in Round 5, and of 50% in Round 6. In each of these rounds, the revenues collected are doubled and redistributed equally to all participants.

Rounds 7 through 9 are more eclectic. In Rounds 7 and 8, we introduce two different progressive taxation schemes. In the first scheme, the top 10% of declared incomes pays a 50% tax rate, the bottom 10% of declared incomes pays a 10% tax rate, and everyone else pays a 30% rate [14]. In the second scheme, all income over 100 ECU is taxed at a 50% rate, income between 50 and 100 ECU is taxed at a 30% rate, and all income below 50 ECU is taxed at a 10% rate. Again, under these two schemes, all taxes collected are doubled and redistributed on an equal basis.

---

[13] Participants have no prior knowledge of the total number of rounds.

[14] However, participants do not know exactly where they themselves fall in the overall distribution of declared incomes.
Finally, in Round 9, we donate all tax revenues (collected under a flat 30% tax rate, and then doubled) to a real world charity, rather than redistributing revenues to the group. The order of experimental scenarios, as well as the rules in each round, are summarized in **S2 Table**.

In each round, participants were informed that they faced an (independent) 5% probability of being audited, in which case those who have under-reported their income must pay a fine equal to twice the amount of uncollected taxes. Importantly, we reveal the results of any audits only at the conclusion of the experiment. Also, at no point during the experiment do participants have information about whether other participants are audited, nor indeed, whether other participants are honestly declaring their own incomes. In fact, we continuously reminded participants at various points through the session that all decisions would be treated anonymously. These procedures were implemented to mitigate the influence of reciprocity, conditional cooperation, reputation or wealth effects.

We also took great care to ensure that the experiment would simulate, as much as possible within a laboratory setting, the private decision problem facing an individual taxpayer. Along these lines, we intentionally incorporate tax language in our protocols, using words such as “income,” “taxes,” and “audit.” While the issue of framing effects in tax experiments is far from settled, we believe this design choice offers an improvement over the use of neutrally-framed compliance games in terms of the ability to stimulate taxpayer motivations.

In summary, by comparing how income is reported across our nine taxation and redistribution scenarios, we are able to investigate differences in tax compliance across a range of parameters. Furthermore, because other researchers have also employed similar experimental designs, we are able to use previous studies as an external check on the validity of our results.

### Results and Discussion

Fig. 1 displays the average percentage of earned income that is reported in each of the nine rounds, broken down between British and Italian participants. The vertical axis displays the average tax compliance rate, defined as the percentage of total earned income that is declared in each round. Several points stand out from the graph. First, comparing Rounds 1 through 3, we see that compliance responds positively to the efficiency of redistribution: in both countries, individuals are more willing to declare a larger percentage of their income when they know that tax revenues produce more public goods. Secondly, individuals respond to higher tax rates by evading their fiscal obligations: compliance falls as we move from Rounds 4 through 6. These results are in line with previous studies, and provide us with some assurance about the validity of our experimental design.

Turning now to our main results, we document a surprising cross-national difference in compliance rates: on average, British participants reported a smaller share of their total income in every round as compared to Italians. As shown in **S2 Fig.**, this finding is also fairly consistent across multiple experimental locations in each country.

While our visual inspection of the the average compliance rate already yields some interesting patterns, such a statistic also hides substantial nuance in participants’ decision-making. Specifically, if we examine the distribution of compliance rates across all reporting decisions, we see that the data are not normal (see Fig. 2). Rather, the average compliance rate actually aggregates three different outcomes:

1. **Complete Compliance:** In over 40% of all decisions, participants declare 100% of their earned income.

---

15 We selected Oxfam for UK participants, and the UNICEF for Italian participants.
Figure 1. Average compliance rates, by country, in rounds 1 through 9. Bar heights represent the average percentage of earned income that is reported. The compliance rate is lower amongst Britons in every single round.

2. **Partial Compliance**: In around 30% of all decisions, participants under-report their income to some degree.

3. **Complete Evasion**: In slightly under 30% of all decisions, participants report that they earned 0 income.

These distributions are similar to the patterns observed in other experimental studies of tax compliance \[22,51\], as well as in “real world” tax declarations. In particular, Alm et al. examined a sample of 1,673 randomly-audited tax returns from the Internal Revenue Service (IRS), focusing solely on self-declared sole-proprietorship income \[51\]. These authors find that the compliance behavior of real taxpayers also conforms to the U-shaped patterns displayed in Figure 2. In other words, the similarity between our results and Alm et al.’s “taxpayer sample” suggests that our participants are treating the experimental decision-problem as a “real life” problem of tax compliance.

Tables 1 and 2 examine how this U-shaped pattern differs between the UK and Italy. Columns (1) through (3) of Table 1 display the proportion of participants who engage in complete evasion. The raw number of participants is also shown in [brackets] next to the proportions. We see that in almost all rounds (except 50% Tax Rate and Charity), a significantly greater percentage of Britons declare exactly 0 income. The UK-Italy gaps are substantively large, ranging from about 8% in Round 7 to almost 18% in Round 2. In columns (4) through (6), we see the corresponding totals for the proportion of individuals who are completely compliant in each round. Here, the data tell a similar story: in the majority of rounds, significantly more Italians report their entire income. The cross-country gaps range from 10% to 18%. Finally, columns (7) through (9) examine the proportion of partial evaders in each country. Here, however, we see that there are no statistically significant differences.

\[16\] This income source was selected because it generates no third-party reporting information (e.g.
Figure 2. Distribution of compliance decisions, all rounds, by country. Bar heights represent the percentage of reporting decisions in which compliance falls within the ranges [0% - 5%], [5% - 10%]...[95% - 100%]. The distribution is predominately bimodal: in around 40% of all decisions, participants declare 100% of their earned income, while in another 30% of all decisions, participants report that they earned 0 income.

However, in addition to considering the raw proportion of partial evaders in each round, we must also examine the percentage of taxes declared amongst partial evaders. We consider this possibility in Table 2. Here, we analyze only the subset of decisions involving partial evasion and ask: “conditional upon engaging in partial evasion, how much income is actually reported?” Interestingly, we detect almost no statistically significant differences between Italian and British participants (with the exception of the very first round). In other words, neither the proportion of partial evaders, nor the behavior of partial evaders (in terms of the percentage of income declared in each round) is very different across our two countries. Rather, it appears that the Italy-UK compliance gap is driven almost entirely by differences in complete compliance and complete evasion.

To what extent can our findings be explained by demographic differences between our participant pools? Controlling for individual-level characteristics that may be correlated with the compliance rate, do Britons still declare less than Italians? To address these questions, we estimate the conditional effect of an Italy country dummy on the compliance rate, holding other factors constant. Following [51], our basic specification is:

\[ Y_{i,t} = \beta_0 + \beta_1 Italy_i + \beta_2 Income_{i,t} + \beta_3 X_i + \psi_t + u_i + \epsilon_{i,t} \]

where the dependent variable \( Y_{i,t} \) denotes the percentage of income reported by participant \( i \) in round \( t \); \( Italy_i \) is a dummy variable for Italian participants; \( Income_{i,t} \) is participant \( i \)’s earned income (standardized) in round \( t \); \( X_i \) denotes a vector of Form W-2 for wage income), and therefore mimics the declaration decision in the laboratory.
Table 1. Proportions of Complete Evaders, Complete Compliers and Partial Evaders: Italy and UK

<table>
<thead>
<tr>
<th>R1: No Redistribution</th>
<th>Complete Evader</th>
<th>(1) Italy</th>
<th>(2) UK</th>
<th>Diff.</th>
<th>Complete Complier</th>
<th>(4) Italy</th>
<th>(5) UK</th>
<th>Diff.</th>
<th>Partial Evader</th>
<th>(7) Italy</th>
<th>(8) UK</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.33 [93]</td>
<td>0.50 [126]</td>
<td>-0.17*</td>
<td>0.33 [93]</td>
<td>0.19 [48]</td>
<td>0.14*</td>
<td></td>
<td>0.34 [95]</td>
<td>0.30 [76]</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>R2: Redistribution</td>
<td></td>
<td>0.26 [73]</td>
<td>0.44 [110]</td>
<td>-0.18*</td>
<td>0.48 [135]</td>
<td>0.31 [77]</td>
<td>0.17*</td>
<td></td>
<td>0.26 [73]</td>
<td>0.25 [63]</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>R3: Redistribution x 2</td>
<td></td>
<td>0.13 [37]</td>
<td>0.30 [76]</td>
<td>-0.17*</td>
<td>0.66 [185]</td>
<td>0.48 [120]</td>
<td>0.18*</td>
<td></td>
<td>0.21 [50]</td>
<td>0.22 [54]</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>R4: 10% Tax Rate</td>
<td></td>
<td>0.16 [44]</td>
<td>0.32 [81]</td>
<td>-0.17*</td>
<td>0.64 [180]</td>
<td>0.48 [119]</td>
<td>0.17*</td>
<td></td>
<td>0.20 [57]</td>
<td>0.20 [50]</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>R5: 30% Tax Rate</td>
<td></td>
<td>0.23 [65]</td>
<td>0.34 [86]</td>
<td>-0.11*</td>
<td>0.47 [131]</td>
<td>0.37 [92]</td>
<td>0.10*</td>
<td></td>
<td>0.30 [85]</td>
<td>0.29 [72]</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>R6: 50% Tax Rate</td>
<td></td>
<td>0.31 [87]</td>
<td>0.38 [95]</td>
<td>-0.07</td>
<td>0.39 [109]</td>
<td>0.35 [87]</td>
<td>0.04</td>
<td></td>
<td>0.30 [85]</td>
<td>0.27 [68]</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>R7: Progressive 1</td>
<td></td>
<td>0.27 [76]</td>
<td>0.35 [87]</td>
<td>-0.08*</td>
<td>0.39 [110]</td>
<td>0.35 [88]</td>
<td>0.04</td>
<td></td>
<td>0.34 [95]</td>
<td>0.30 [75]</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>R8: Progressive 2</td>
<td></td>
<td>0.21 [59]</td>
<td>0.32 [80]</td>
<td>-0.11*</td>
<td>0.39 [110]</td>
<td>0.29 [73]</td>
<td>0.10*</td>
<td></td>
<td>0.40 [112]</td>
<td>0.39 [97]</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>R9: Charity</td>
<td></td>
<td>0.14 [38]</td>
<td>0.17 [42]</td>
<td>-0.03</td>
<td>0.69 [195]</td>
<td>0.63 [158]</td>
<td>0.06</td>
<td></td>
<td>0.17 [48]</td>
<td>0.02 [50]</td>
<td>-0.03</td>
<td></td>
</tr>
</tbody>
</table>

The raw number of participants is shown in brackets, next to the proportion. We employed Schlag’s Z-test to test for country-level differences in columns (3), (6) and (9). * indicates whether differences between countries are statistically significant at the 5% level.

Table 2. Compliance Rate of Partial Evaders: Italy and UK

<table>
<thead>
<tr>
<th>R1: No Redistribution</th>
<th>Number of Partial Evaders</th>
<th>(1) Italy</th>
<th>(2) UK</th>
<th>% of Income Declared by Partial Evaders</th>
<th>(3) Italy</th>
<th>(4) UK</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>95</td>
<td>76</td>
<td>0.59</td>
<td>0.47</td>
<td>0.11*</td>
<td></td>
</tr>
<tr>
<td>R2: Redistribution</td>
<td></td>
<td>73</td>
<td>63</td>
<td>0.52</td>
<td>0.48</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>R3: Redistribution x 2</td>
<td></td>
<td>59</td>
<td>54</td>
<td>0.51</td>
<td>0.55</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>R4: 10% Tax Rate</td>
<td></td>
<td>57</td>
<td>50</td>
<td>0.60</td>
<td>0.49</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>R5: 30% Tax Rate</td>
<td></td>
<td>85</td>
<td>72</td>
<td>0.55</td>
<td>0.49</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>R6: 50% Tax Rate</td>
<td></td>
<td>85</td>
<td>68</td>
<td>0.51</td>
<td>0.48</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>R7: Progressive 1</td>
<td></td>
<td>95</td>
<td>75</td>
<td>0.50</td>
<td>0.47</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>R8: Progressive 2</td>
<td></td>
<td>112</td>
<td>97</td>
<td>0.51</td>
<td>0.49</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>R9: Charity</td>
<td></td>
<td>48</td>
<td>50</td>
<td>0.55</td>
<td>0.53</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

We employed Mann-Whitney tests to test for country-level differences in column (5). * indicates whether differences between countries are statistically significant at the 5% level.

demographic variables including gender, age, employment status, economics training, previous participation in behavioral experiments, self-reported risk attitudes, and self-reported beliefs about the behavior of other participants; \( \psi_t \) is a set of \( T - 1 \) dummies that capture round fixed effects; \( u_i \) are random effects that control for unobservable individual characteristics; and \( \epsilon_{i,t} \) is an individual-round error term. Following [51], we report results for a (participant) random effects generalized least squares estimation with standard errors corrected for clustering at the individual level. Results are presented in Table 3.

Table 3 shows that gender, economics training, past participation in experiments, the number of observations drops slightly once we include demographic covariates in model 3 of Table 3. This is because in early versions of the experiment, the experimental tasks were implemented in zTree, while the demographic information was collected separately using Qualtrics survey software. This necessitated participants entering their anonymous participant-IDs twice: once into zTree, and once again into Qualtrics. Because some participants accidentally entered different participant-IDs into the two systems, we were unable to match their experimental decisions with their demographic data. Later in our project, we were able to integrate the experimental and survey portions of the study.
Table 3. Estimates of the Compliance Rate

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>First-time Particip.</th>
<th>Repeat Particip.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Italy</td>
<td>0.13**</td>
<td>0.12**</td>
<td>0.10**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Income (standardized)</td>
<td>-0.03**</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Male</td>
<td>-0.17**</td>
<td>-0.05</td>
<td>-0.19**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Age (standardized)</td>
<td>0.01</td>
<td>0.07**</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Employed</td>
<td>-0.04</td>
<td>0.04</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.06)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Economics Training</td>
<td>-0.09**</td>
<td>-0.10</td>
<td>-0.08**</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Previous Participation</td>
<td>-0.09**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk (standardized)</td>
<td>-0.06**</td>
<td>-0.07**</td>
<td>-0.06**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Others Report: “Less”</td>
<td>-0.12**</td>
<td>-0.18**</td>
<td>-0.11*</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.06)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Others Report: “Much Less”</td>
<td>-0.26**</td>
<td>-0.35**</td>
<td>-0.25**</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.08)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.52**</td>
<td>0.37**</td>
<td>0.74**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Wald χ²</td>
<td>20.94**</td>
<td>460.2**</td>
<td>871.8**</td>
</tr>
<tr>
<td>Number of Participants</td>
<td>531</td>
<td>531</td>
<td>512</td>
</tr>
<tr>
<td>Number of Decisions</td>
<td>4779</td>
<td>4779</td>
<td>4608</td>
</tr>
<tr>
<td>Round Fixed Effects</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Panel estimations with participant random effects and clustered (participant level) standard errors. The dependent variable is the percentage of earned income that is declared for tax purposes by individual \(i\) in round \(t\). * and ** indicate significance at the 5% and 1% levels, respectively.

Risk attitudes and beliefs about others’ behavior are all correlated with compliance.  

---

**18** We measure risk-attitudes with a survey item which asks participants to rank themselves on a 10-point scale, with 1 signifying a person who “normally tries to avoid taking risks” and 10 signifying someone who is “completely willing to take risks.” The question is very similar to one asked on the German Socio-Economic Panel (SOEP). While simple to understand, this method presents a drawback in that individuals’ answers are not incentivized. While we appreciate this concern, we would point to two counterarguments supporting the validity of our measure. First, even though risk attitudes elicited in this manner might constitute a form of “cheap talk,” there is no reason to suspect that individuals would intentionally misrepresent their risk preferences: nothing can be gained (either in terms of money, reputation or self-image) from expressing a “low” or “high” appetite for risk. Secondly, Dohmen et al. conducted a “horse race” between different risk elicitation methods (both incentivized and non-incentivized) in terms of the ability to explain behaviors such as holding stocks, occupational choices, and smoking, and find that the general risk question “generates the best all-round predictor of risk behavior” [52]. In the end, we agree with Charness et al. that the elicitation approach should be tailored to the underlying question to which the research is addressed [53]. Since our aim is solely to measure differences across individuals in risk preferences (as opposed to, for example, estimating specific parameters of a mathematical model), we opt for the simplest method. We believe that the gains from having an easily understandable question outweigh the risks of “cheap talk.”

**19** We measure beliefs by simply asking participants: “Do you think most of the participants in the
Several results are worth emphasizing: first, unsurprisingly, participants who are more risk-taking also comply less\(^{20}\). Second, individuals who think that others are dishonest are also less likely to comply themselves. This echoes other findings on the importance of social norms for tax compliance\(^ {55}\). Finally, in line with earlier studies\(^ {56-60}\), we also show that men and economists are less compliant\(^ {22}\).

Most importantly, we find that even controlling for all of these characteristics, the Italian country-dummy remains substantively strong and statistically significant. While a simple comparison of the average compliance rate across countries (model 1) shows that Italians report about 13% more of their total income than British participants, this estimate falls only slightly once we include round fixed effects and a control for earned incomes (model 2), as well as other individual level controls (model 3). Importantly, the core finding that Italian participants are systematically more compliant in their fiscal decision-making remains robust in our regression models\(^ {23}\).

To what extent can we interpret these results as evidence that Italians possess a stronger ethics of tax compliance? Are “culturalist” arguments about the (lack of) morality amongst southern European taxpayers simply wrong? Or might other factors that we have not controlled for account for our surprising results? Here, we discuss two potential confounds: trust in anonymity, and group-level reputational concerns.

First, although we emphasize the anonymity of decisions at multiple points throughout the experiment, the propensity of participants to trust our assurances may vary across countries. For instance, Italian participants may be more concerned that their anonymity could somehow be compromised, and therefore act more compliantly than they would otherwise.

While we cannot directly control for how “safe” participants felt about the confidentiality of their decisions, we can test one implication of this argument. In particular, we may reasonably assume that first-time participants would feel the most insecure since, presumably, individuals who have taken part in several experiments are more familiar with anonymity procedures. Therefore, if trust in confidentiality did vary cross-nationally in the way we describe above, then we might expect that the country-level compliance gap would be especially large amongst first-time participants. In fact, as shown in models (4) and (5) of Table 3 we find the opposite: amongst first-timers (i.e. those least likely to believe that anonymity would be preserved) the coefficient on Italy\(_i\) is substantively smaller (7% vs. 10% in model 3) and statistically insignificant. By contrast, amongst repeat participants, the Italy dummy is estimated to be 11%, which is slightly larger than the effect we identified in model (3). These results suggest that our findings are not due to differences in the degree to which participants are fearful that their anonymity would be compromised.

A second potential confound relates to a concern we have already noted about maintaining a positive reputation. Specifically, participants in the lab know that they are being observed, and this feeling of being “under the microscope” (especially with experiment reported (a) their total earnings, (b) less than their total earnings, or (c) much less than their total earnings for tax purposes?) While recognizing the arguments in favor of more complicated and incentivized belief elicitation methods\(^ {54}\), we opted again for a simple survey measure in order to minimize comprehension issues.

\(^{20}\) On average, British participants self-rate as more risk-taking. Thus, the inclusion of the risk variable in the regression with only the Italy\(_i\) dummy tends to shrink the cross-country compliance gap from 13% to 8% (results not shown). However, the coefficient on Italy\(_i\) is still significant at the 1% level.

\(^{21}\) Here, we note that Italians are on average more pessimistic about the behavior of their counterparts. Thus, controlling for beliefs in the regression slightly increases the size of the cross-country gap from 13% to 14% (results not shown).

\(^{22}\) Since there are slightly more economists amongst Italian participants (as shown in S1 Table), controlling for this variable also inflates the cross-country compliance gap from 13% to 15% (results not shown).

\(^{23}\) We also experimented with a series of tobit models, but our findings remain substantively the same. The results of these regressions are available from the authors.
respect to a sensitive topic such as tax evasion) may bring about more compliance than would otherwise prevail in the “real world.” Furthermore, the size of this effect may differ between Italy and the UK. In particular, Italians may be more concerned about their collective reputation as a “high evasion” nation. They may therefore act more compliantly in the experiment in order to show their rejection of this stereotype. Importantly, this dynamic may still obtain even though (a) our moderators are all native speakers, and (b) participants are not informed that they would be compared to individuals in another country.\(^{24}\)

While we cannot rule out this alternative explanation for our results, we note that this dynamic is also present in other cross-national experimental studies on tax compliance.\(^{14,36}\) Moreover, in Lewis et al.’s survey experiment directly comparing Italian and British participants, reputational concerns did not seem to prevent Italians from under-declaring their incomes,\(^{61}\) a telling result given that the declarations in the Lewis study were purely hypothetical, and thus the “cost” of maintaining a positive reputation was lower than in our experiment. Therefore, although we acknowledge that reputation concerns may be potential confounding variable, we believe that these concerns do not play a prominent role in accounting for our surprising findings.

Conclusions and Implications

Turning now to the original question animating our study, what light can our results shed on the current fiscal crises facing European countries? Clearly, many northern Europeans (and Americans) hold the view that the difficulties facing Greece, Italy, Spain and others are not simply the product of poor institutions or badly designed public policies, but are instead symptomatic of a far more fundamental problem – one of culture.\(^{25}\) In a nutshell, southern Europe’s fiscal conundrum is said to stem from the fact that, in such “limited morality” societies, the “moral costs” of non-compliance are insufficient to ensure effective tax collection.

The problem, analytically, is that culture and institutions are difficult to disentangle and are almost certainly interdependent. For example, Italians may readily cheat on their taxes in “real life” precisely because they believe that the state - in corruptly and inefficiently using their tax revenue - is also cheating them. In fact, recent public opinion polls have shown that southern Europeans give their governments consistently low scores on control of corruption and the quality of service delivery.\(^{62,63}\) By contrast, northern Europeans’ high willingness to pay may be a direct reflection of their belief that taxes go towards important public services which they value and personally consume.

In this article we have tried to analyze the specific question of whether southern Europeans (in this case, Italians) would behave differently than northern Europeans (in this case, Britons) when faced with exactly the same institutions. We discovered, much to our own surprise, that British participants are more likely to under-declare their incomes in a tax / public goods experiments than Italian participants. While we do not pretend that our study has isolated all of the cultural variables that may influence tax compliance decision-making, at a minimum, our results cast doubt on the above-mentioned “culturalist” arguments.

We find these results encouraging. If it were indeed the case that cultural variables dominate institutional structure, then the prospects for Europe – and for much of the developing world today – would be grim indeed. Instead, we show that, when given the opportunity to contribute to and share in collective goods on an equal institutional

\(^{24}\)We are grateful to an anonymous reviewer for bringing our attention to this point.

\(^{25}\)See, for example, Alan Greenspan’s editorial “Europe’s crisis is all about the north-south split” at [http://on.ft.com/nZGC15](http://on.ft.com/nZGC15).
playing field, Italians behave no worse than their British counterparts. Thus, we believe that a specific focus on institutional reforms and improving the “Quality of Government” [24] are likely to yield more significant results than the cultural blame game that too often seeps into the policy debate. Rather than the syndrome of southern European “amoral familism,” our results suggest that more powerful explanations for national variation in tax compliance are rooted in differences in institutions.

Acknowledgments

We thank Paul Bauer, John D’Attoma, Fillipos Ekadaktylos, Jonas Fooken, Johanna Gereke, Benedikt Hermann, Massimo Motta, Martijn Schoonvelde and two anonymous reviewers for helpful comments on this paper.

Author Contributions

Conceived and designed the experiments: GA SO FP SS. Performed the experiments: GA SO FP SS NZ. Analyzed the data: NZ. Wrote the paper: NZ.

References


Supporting Information

S1 Fig.
Example screenshot: Earnings reporting screen.

S2 Fig.
Compliance rate across all locations.

S1 Table
Summary of participant characteristics: Italy and the UK.

S2 Table
Summary of tax reporting rounds.

S1 Text
English Language Experimental Instructions.

S2 Text
English Language Questionnaire.

S3 Text
IRB Approval Letters.