Battiston, P., Gamba, S. (2016)
The Impact of Social Pressure on Tax Compliance: a Field Experiment
International Review of Law and Economics, 46, 78–85

The Impact of Social Pressure on Tax Compliance: a Field Experiment

Pietro Battiston[a,b,*], Simona Gamba[c,d]

[a] *DEFAP, Department of Economics, Management and Statistics, University of Milano Bicocca, Piazza Ateneo Nuovo, 1, 20126 Milano, Italy.*
[b] *Present address: Istituto di Economia, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, 56127 Pisa, Italy.*
[c] *DEFAP, Department of Economics, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123 Milano, Italy.*
[d] *Present address: FBK-IRVAPP, via Santa Croce, 77, 38122 Trento, Italy.*

Abstract

We study the effect of social pressure on tax compliance, focusing on the compliance of shop sellers to the legal obligation of releasing tax receipts for each sale. We carry out a field experiment on bakeries in Italy, where a strong gap exists between the legal obligation and the actual behavior of sellers. Social pressure is manipulated by means of an explicit request for a receipt when not released. We employ an innovative approach to the identification of the treatment effect. We find that a single request for a receipt causes a 17 per cent rise in the probability of a receipt being released for a sale occurring shortly thereafter, causing on average more than two receipts to be released. We also find strong evidence of persistence in compliance decisions.

JEL classification: C93, H32, K34

Keywords: Tax evasion, field experiment, peer pressure, social pressure.

*Corresponding author.
Email addresses: `me@pietrobattiston.it` (Pietro Battiston), `gamba.simona@gmail.com` (Simona Gamba)

1. Introduction

The literature on fiscal compliance has developed from the seminal model of Allingham and Sandmo (1972). In their work (as in similar studies by Kolm, 1973, and Singh, 1973), expected utility maximizing agents choose the income level to be reported to the fiscal authority, considering the probability of being audited and the size of the fine. At the empirical level, however, researchers have faced a major puzzle: in all advanced economies, the level of tax compliance is far higher than the one predicted by such theories (Graetz and Wilde, 1985, Alm et al., 1992). This paper deals with such puzzle by proposing and implementing a new experimental design.

A stream of literature has approached the discrepancy by extending the original model with more realistic specifications of the context in which tax declaration decisions are taken. In this context, financial strain (Wärneryd and Walerud, 1982) and the broad category of *opportunities* have been analyzed, among other factors. The role of third-party reporting, which limits the possibility for employees to evade taxes, has been widely discussed (Andreoni et al., 1998) and tested experimentally (Slemrod, 2007 and Kleven et al., 2011). Even these studies, however, recognize that the high level of compliance which is observed empirically cannot be fully explained without taking into account behavioral factors. This view, which is nowadays widespread in the literature, is hence the starting point for the present study.

Among the several authors who extended the basic model by Allingham and Sandmo with the inclusion of non-monetary motives, Bordignon (1993) embeds fairness-based evaluations into the utility function, while Gordon (1989) introduces non-pecuniary *stigma* costs associated with tax evasion. Weigel et al. (1987) and Groenland and Van Veldhoven (1983) provide a *social and psychological model*, which represents a broader approach to the several conditions which influence fiscal behavior, such as personality (Lewis, 2011). Studies on behavioral aspects of tax compliance are rooted in the wider stream of literature about the social aspects of deterrence (see for instance Grasmick and Bursik Jr, 1990 and Paternoster et al., 1983).

We contribute to such literature by testing, through a field experiment ran in shops, the salience of *direct peer pressure* for the fiscal compliance of *sellers*. There are multiple channels through which peer pressure can influence the fiscal behavior; for ease of exposition, we regroup them in three classes: *honesty*, *opportunity*, and *conformism*. Concerning the first class, a vast empirical literature points at the importance of *social norms* in regulating human behavior: peer pressure can signal the status of fiscal compliance as a social norm, pushing the taxpayer who wants to avoid the psychic cost related to fiscal evasion towards honesty. The opportunistic channel consists in the possibility that, for instance, a seller who is not personally intimately concerned with the social norm shifts her behavior towards compliance, in order not to loose the part of her customer base which favors honesty. Finally, because of conformism, the mere knowledge, or feeling, that fiscal compliance is the typical behavior in the community of reference could both raise the perceived probability of audits (i.e. by representing a signal of the probability perceived by others) and stimulate conditional cooperation (Fellner et al., 2013). Empirically disentangling such effects is a difficult task, but a consistent literature has identified the first as particularly relevant for the taxpayer decision (see Kirchler, 2007 pp. 64-65, and Galbiati and Zanella, 2012), while Erard and Feinstein (1994) merge the approaches of tax morale and utility maximization by showing that in presence of a subpopulation of *honest* taxpayers, even purely selfish citizens end up paying more taxes. The goal of our experiment is not to isolate one specific channel, but rather to measure empirically the overall effect that peer pressure has on the compliance choice, a measure which has relevant policy implications.

Studies of tax compliance have been historically confronted with a lack of data that is particularly hard to overcome, as effectively summarized by Cowell (1991): *"Data from official investigations are hardly ever available and data from other sources may be suspect: if you could directly observe and measure a hidden activity, then presumably it could not really have been properly hidden in the first place."* Weigel et al. (1987) considered as fundamental for future fiscal research the development of *creative methods* for attaining objective estimates of tax evasion behavior. The quest is still open, as reported more recently by Halla (2012). In particular, the frequent use of survey data, where individuals self-report their tax behavior, has since long been perceived as a crucial issue (Weigel et al., 1987, Elffers et al., 1987), because of the possible misreporting.

Therefore, a growing stream of literature has focused on *experiments* aimed at reproducing the economic and psychological reasoning behind tax compliance. This stream can be traced back to Reis and Gruzen (1976) and Kidder et al. (1977); more recent attempts in this direction are those of Alm et al. (1992) and Cummings et al. (2006). In his exhaustive review of the field, Torgler (2002) acknowledges the relevance of experiments in that tax enforcement, tax rate and income levels can be controlled.

The effect of social norms and social disapproval on tax compliance has been approached experimentally for instance by Bosco and Mittone (1997). Their design allows to test two separate hypotheses, concerning the effects of either *subjective* or *collective* moral constraints on tax compliance. Subjective moral constraints are manipulated as follows: while in the control group money collected through taxes is just taken away from participants, in the experimental treatment there is a partial redistribution of the collected amount. In order to test for the second hypothesis, instead, a treatment is run in which the identity of individuals who are caught cheating is publicly revealed, and evaders hence run the risk of being identified as such by other participants. The authors find significant evidence only in favor of the first hypothesis. More recent examples of experimental studies on the effect of social pressure on the compliance choice can be found in the work of Cummings et al. (2001), Alm et al. (2007), and Fortin et al. (2007).

However, still Torgler (2002) casts doubts on the fact that laboratory experiments can be considered informative about actual tax compliance behavior. This concern is shared by Halla (2012), who suggests that individuals react to experimenters' stimuli differently than with real tax authorities. Indeed, social norms are part of the *culture* of any society, of which a laboratory experiment allows to study only schematized traits, and at the same time they are a fundamental ingredient of the compliance decision (Posner, 2000) because they *"constitute constraints on individual behavior beyond the legal, information and budget constraints usually considered by economists"* (Fehr et al., 2002).

Although their number has been recently increasing, relatively few attempts have been made to identify the size and the determinants of tax evasion through the use of field experiments. This is due in part to the typical reluctance of national fiscal authorities toward randomized actions (other than budget-motivated randomized audits such as those described by Erard et al., 2002), which are supposed to go against the principle of equity.[1] The approach of Schwartz and Orleans (1967), later adopted by Wenzel (2001), is based on surveys sent to taxpayers some time before they file their tax declaration. The questions asked vary from group to group: this enables the

[1]Randomized setups are characterized precisely by the fact that they treat equal citizens differently, rather than shaping enforcement actions deterministically on observable variables.

authors to identify the reduction in evasion due to *"conscience"* versus the one due to *"sanctions"*, by arousing respectively the feeling of *guilt* related to the social loss or the *fear* of detection. While they find that the relative importance of the two motives depends on the social and economic status of individuals, overall they report that *"conscience appeals are more effective than sanction threats"*. Slemrod et al. (2001), through threat-of-audit letters, identify the response of taxpayers to an increase in audit probability, and report mixed evidence. They find an increase in amounts declared by low and middle-income taxpayers, but a *decrease* in amounts declared by high-income ones. This result is attributed to the particular wording used in the letters, together with the heterogeneity of beliefs and of information that individuals have about the fiscal authority. Kleven et al. (2011) bring into the picture the effects of an audit itself on subsequent tax declarations, as an indicator of undeclared income. Their main conclusion is that fiscal evasion is severely hindered by third-party reporting. Still, they acknowledge the evidence of behavioral factors: even though audits do not imply a higher audit probability in the future, they have a positive deterrence effect for the following fiscal year. Finally, Fellner et al. (2013), in addition to independently testing the effect of a *threat* (a message directed at changing the perceived sanction risk) and of a *moral appeal* (stressing that evasion is an act against *fairness*, which harms honest taxpayers), introduce the innovative element of *social information*. A subsample of their subjects is informed of the compliance rate for the specific TV license fees on which the experiment is based. The authors show that the effect of such new information goes in the direction of *conformity*, by increasing (decreasing) the compliance of individual with lower (higher) prior expectations on the compliance rate.

While Fellner and coauthors interestingly bring into the picture the effect of social pressure, in their study, as in the other field experiments previously cited, the treatment comes from the interaction of citizens with *institutions* - in particular, it is determined in the context of the surveys, audits, or threat-of-audit letter that these institutions implement. Instead, to the best of our knowledge, no field experiment on fiscal compliance has been previously implemented focusing on the *direct* effect of social pressure between peers, as in the tradition of experiments on peer pressure started by the seminal work of Asch (1955) (also see Falk and Fischbacher, 2002 and Falk and Ichino, 2005). The present paper tries to fill this gap. It does so by exploiting the particular case of tax evasion among shop sellers in Italy, a country where non-compliance is relatively widespread (as confirmed both by official reports, and by our experimental data). The vast majority of Italian shops are obliged by the law to release a tax receipt for each sale (supermarkets and kiosks are among the few exceptions, which however are irrelevant in the present context). The total sum of receipt amounts represents the revenues of a shop, and receipts themselves constitute a proof for the fiscal authority. Both value added tax and income tax are then calculated on the basis of such revenues. As a consequence, the omitted release of a receipt is an act of fiscal evasion (and is in principle punished as such by the law), allowing the seller to evade both categories of taxes. Interestingly, this act of tax evasion is not only common, but also, at least in the case under analysis, committed *openly*, making it trivial for a purchaser to ascertain non-compliance. Although it would also be trivial for the purchaser to actively *fight* tax evasion - by simply requesting the receipt when it is not released - this behavior is far from being widespread. As will be clarified later, when such a request is made, it is an unambiguous act of disinterested social pressure.

The remainder of the paper is structured as follows. We first describe our experimental design in Section 2. Section 3 presents the results, which are discussed in Section 4. Finally, Section 5 summarizes our conclusions.

2. Experimental design

The aim of the experiment is to study the effect of peer pressure on tax compliance, focusing on the compliance of shop sellers to the legal obligation of releasing tax receipts for each sale.[2] The treatment is the request of a receipt not spontaneously released.

Fiscal compliance in Italy is well known to vary from city to city: in order to get a meaningful estimate, we focused on bakeries in the central areas of Milan. The experimental sample consisted of 108 bakeries: for each bakery, the time line of the experiment was articulated in two periods. In period 1, an agent entered the shop and bought a loaf of bread. If the receipt was not released, the agent would ask for it:[3] this request was our treatment.

[2] The Italian law dictates that a receipt is printed for each sale, and contextually released to the buyer. The fine for not releasing a receipt is of 129 € *for any purchase totaling less than 516 €*, and five violations in a time span of five years, *independently from the entity of the sales*, result in a suspension of the commercial activity for 15 days.

[3] The receipt was always requested using the same wording ("*Vorrebbe essere così gentile da rilasciarmi lo scontrino?*", which roughly translates to "*Would you be so kind as to*

In period 2, twelve minutes after the first agent left the shop, another agent entered the same bakery. Following the same procedure as in period 1, the agent bought a loaf of bread of a different type. The role of this second agent was to assess if the receipt was now given. Whatever was the behavior of the seller, *no request for a receipt took place at this time.*

In all cases, the purchase was paid with an amount of money higher than its cost,[4] so that the agent had to wait for the change. This design choice was made because a client standing still after having paid and received the bread would have probably influenced the behavior of the seller. In this way instead, the moment in which the change was given (with or without the receipt) represented unambiguously the end of the transaction. The choice of the twelve minutes time span was made because it is absolutely unlikely that any client would spend such an amount of time in a bakery. This means that when the second agent entered, the first one, as well as the clients present in the shop when the request for the receipt had taken place, had already left the shop. In this way, any change in the behavior of sellers can be attributed uniquely to a reaction of the sellers themselves to the request, as opposed to *indirect* pressure, or to the presence of the client who proved to be particularly "picky".

Several measures were adopted to ensure that the two passes had on average the same exact characteristics, except for the treatment (if any). In particular, (a) the entry order of the two agents (one male and one female, both around 25 years of age) was randomized; (b) the types of bread purchased were randomized,[5] and most importantly (c), the second agent *did not know* if the first had been spontaneously given the receipt and hence if the bakery had been treated.

The particular category of businesses which we study - bakeries - exhibits several features that make it particularly suitable for our experiment. First, the good at sale, bread, is relatively standardized, making it meaningful to compare different shops. Second, it has a low cost, which implies that the

---

give me the receipt?").

[4]For the sake of homogeneity, banknotes were never used, and the amount given was always lower than 2 €.

[5]Each time, one agent asked for a type of bread and the other one asked for another, resorting to a third and then to other types if the requested one was not available. The three types chosen are comparable in weight, size, cost, and all of them are usually sold by any bakery.

profit obtained by evading is generally not the object of bargaining between the seller and the buyer.[6] Therefore, the act of requesting the receipt does not affect the utility of the buyer: it can instead be considered as a disinterested act of social pressure.

We employ an innovative empirical approach **(Figure 1)** which allows us to increase the power of the statistical tests without affecting their validity and interpretation. Let Y_1 (respectively Y_2) be the decision to release the receipt at the first (second) pass, and D a boolean variable indicating the selection for the treatment. After randomly determining the subsample of bakeries for which $D = 1$, it would be possible to estimate the effect of the treatment on the probability of switching from non-compliance to compliance. In the terminology of the typical treatment-effect framework, this would correspond to an Average Treatment effect on the Treated (i.e. on non- compliant bakeries):

$$\begin{aligned} ATT =&\mathbb{P}\{Y_2 = 1|Y_1 = 0, D = 1\} - \mathbb{P}\{Y_2 = 1|Y_1 = 0, D = 0\} \\ =&\frac{\mathbb{P}\{Y_1 = 0, Y_2 = 1, D = 1\}}{\mathbb{P}\{Y_1 = 0, D = 1\}} - \frac{\mathbb{P}\{Y_1 = 0, Y_2 = 1, D = 0\}}{\mathbb{P}\{Y_1 = 0, D = 0\}}. \end{aligned} \quad (1)$$

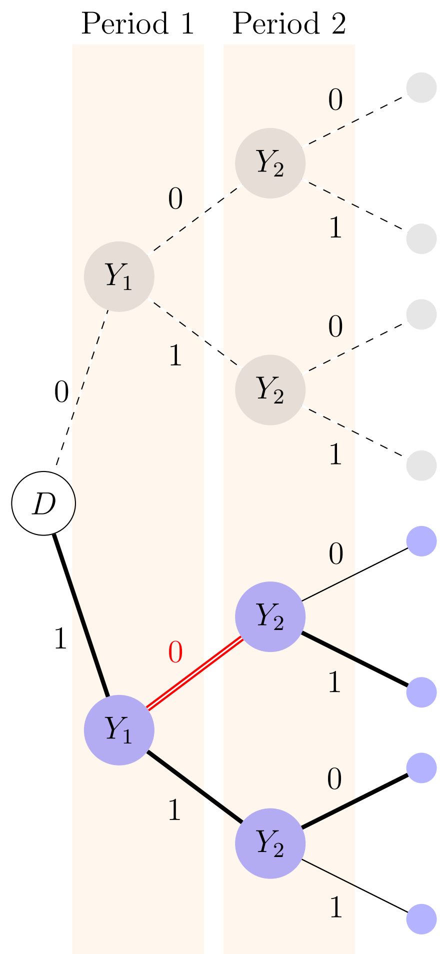
Our experiment is peculiar in the fact that the treatment is exerted only for subjects for which $Y_1 = 1$, and only *after* Y_1 is realized **(see double link in Figure 1)**: hence it cannot influence Y_1. This means that **the two denominators in Equation 1 coincide.** Moreover, we can rely on the following identifying restriction.

> **Exchangeability assumption**: in absence of treatment, both possible switches in the compliance decision happen with the same frequency:
>
> $$\mathbb{P}\{Y_1 = 0, Y_2 = 1|D = 0\} = \mathbb{P}\{Y_1 = 1, Y_2 = 0|D = 0\}.$$

This assumption holds because it resorts to assuming that fiscal compliance does not significantly increase in a time span of 12 minutes. Together with the fact that $\mathbb{P}\{Y_1 = 1, Y_2 = 0|D = 0\} = \mathbb{P}\{Y_1 = 1, Y_2 = 0|D = 1\}$

[6] Bargaining is known to take place in other sectors, in which the amount of evasion benefits *per purchase* is much higher and the buyer is often offered a discount conditional on not receiving the receipt.

Figure 1: **Decision tree describing the experimental setup**



Note: **the whole tree describes the "traditional" setup: D denotes the selection into treatment, Y_1 and Y_2 the decision to release a receipt for the first and second sale, respectively. The treatment is implemented in correspondence of the red link (hence, only in the subtree $\{D = 1, Y_1 = 0\}$). In our case, setting $D \equiv 1$ (i.e. omitting the dashed subtree) can be done without affecting the interpretation of the results, which ultimately come from the comparison of the two thick paths. This approach results in an increased sample size in the bottom subtree. See text for details.**

(since by construction, no one with $Y_1 = 1$ is asked to release the receipt), this implies that observations which would be dropped according to the **"traditional"** approach (i.e. when $Y_1 = 1$) provide instead the control group. Formally, **Equation** (1) **can be reformulated** as:

$$ATT = \frac{\mathbb{P}\{Y_1 = 0, Y_2 = 1, D = 1\} - \mathbb{P}\{Y_1 = 1, Y_2 = 0, D = 1\}}{\mathbb{P}\{Y_1 = 0, D = 1\}}. \tag{2}$$

Notice that Equation (2) can be estimated restricting to the subsample for which $D = 1$ **(bottom subtree in Figure 1)**. This means we can effectively refrain from defining ex ante a control group, and set $D \equiv 1$ for all observations (treating all non-compliant sellers, rather than randomizing the treatment). Thus, we obtain:

$$ATT = \frac{\mathbb{P}\{Y_1 = 0, Y_2 = 1\} - \mathbb{P}\{Y_1 = 1, Y_2 = 0\}}{\mathbb{P}\{Y_1 = 0\}} \tag{3}$$

The fact that the effective treatment status of a given observation is not determined ex ante is shared with other papers in the literature (e.g. Levitt and Wolfram, 1997). Instead, the use of the exchangeability assumption to provide a counterfactual is, to the best of our knowledge, an original contribution of the present paper. It is made possible by the fact that groups $\{Y_1 = 0, Y_2 = 1\}$ and $\{Y_1 = 1, Y_2 = 0\}$ are ex ante identical. Our experimental design allows to observe $\mathbb{P}\{Y_1 = 0, Y_2 = 1\}$, $\mathbb{P}\{Y_1 = 1, Y_2 = 0\}$ and $\mathbb{P}\{Y_1 = 0\}$, and hence to **accomplish the same objective as the "traditional" approach, i.e. the estimation of the ATT, which answers** the question **"to what extent** does exerting social pressure *on non-abiding sellers* affect their propensity to tax compliance?".[7]

In our approach, *all* shops such that $Y_1 = 0$ are treated, rather than a random subset of them **(those such that** $D = 1$**)**: the total number of treated subjects hence **increases (e.g. it doubles, compared to the**

[7] Notice that it would be hardly interesting to measure an ATE (Average Treatment Effect) - that is, to treat bakeries where the receipt is spontaneously given. In principle, an experiment could be ran in which social pressure is exerted *at the start* of the transaction, for instance with agents explicitly stating, at the moment of asking the loaf of bread, that they want the receipt, but this was not our choice for two reasons. First, the measurable effects would have been largely diluted. Second, it would be suspicious if a client asked for the receipt beforehand.

case in which half of the sample is assigned $D = 0$). Given the relative scarcity of non-compliant bakeries[8] (see Section 3), this implies an increase of the power of a statistical test of the ATT, without interfering with the causal interpretation.

For the interpretation of the results, two aspects of our experimental design are worth stressing. First, although explicit requests for receipts are presumably rare in Italy, all other aspects of the interaction between the buyers and the sellers in the experiment are absolutely ordinary. The change of agent and of type of bread being requested from one pass to the other, together with the fact that bakeries are characterized by a high number of low volume sales, make it virtually impossible that any seller noticed anything unusual - apart obviously from the rebuke, when there was one. Second, although the seller can have, as already mentioned, an opportunistic response, the request for the receipt itself can *only* be interpreted as an act of gratuitous social pressure, signaling adherence to the social norm of fiscal compliance, rather than self-interest. Bread is not covered by any warranty for which the receipt could serve as a proof of the purchase, the Italian legislation does not envisage sanctions for clients unable to show the receipt of a purchase just made,[9] and although in principle a client can denounce a seller for not releasing a receipt, this becomes impossible precisely after the receipt has been (requested and) released.[10]

The experiment was ran in the first 2 weeks of March 2012 and involved 108 bakeries**: 38 which had not released the receipt during a previous survey,**[11] **and 70 others located in their proximity.** The peculiarity of the sample is taken into account in Section 3, **where we consider how the**

[8]The method we adopt can be applied in different contexts, and the share of subjects for which $D = 0$ will not necessarily be equal to 0 or 1: the appropriate choice will depend on the natural frequency of $Y_0 = 1$, and will be 0 as long as such natural frequency is below $\frac{1}{2}$, as in our case.

[9]Such sanctions were theoretically present, although very rarely implemented, until 2003, when a legislative change left only the existing sanctions on sellers.

[10]On the relationship between requests and tax evasion, also see Fabbri and Hemels (2013).

[11] In order to obtain a preliminary assessment of tax compliance **and to prepare the experimental sample (non-compliant sellers, which are central to our design, are relatively scarce),** a single pass was carried out on 177 bakeries in January 2012**. During this pass, in which no treatment was implemented,** a non-compliance rate of 22% was observed.

possible correlation between the propensity to evade and the effect of the treatment can affect our estimates. During the experiment, 21% of bakeries did not release the receipt at the first pass, and were hence treated. Of these, 13 were treated by a female agent, 10 by a male agent (each agent entered as first in exactly 50% of the bakeries). Among the treated bakeries, one type of bread had been asked in 11 cases, and the other type in 12 cases.

3. Results

Table 1 summarizes the main features of the experimental data. Notice that the overall compliance rate during the second pass ($Y_2 = 1$) was 82.4%, higher than during the first pass (78.7%).

Table 1: Summary of experimental data

		Second pass (Y_2)		
		0	1	Total
First pass (Y_1)	0	7 (6.5 %)	16 (14.8 %)	23 (21.3 %)
	1	12 (11.1 %)	73 (67.6 %)	85 (78.7%)
	Total	19 (17.6%)	89 (82.4%)	108 (100%)

Note: number of shops observed in each of the four possible combinations of compliance at first and second pass; 0 refers to a receipt not *spontaneously* released, 1 to a receipt regularly released. Last row and last column: partial sums, and relative frequencies. Values in row $Y_1 = 0$ correspond to *treated* bakeries.

By estimating Equation (3) we obtain:

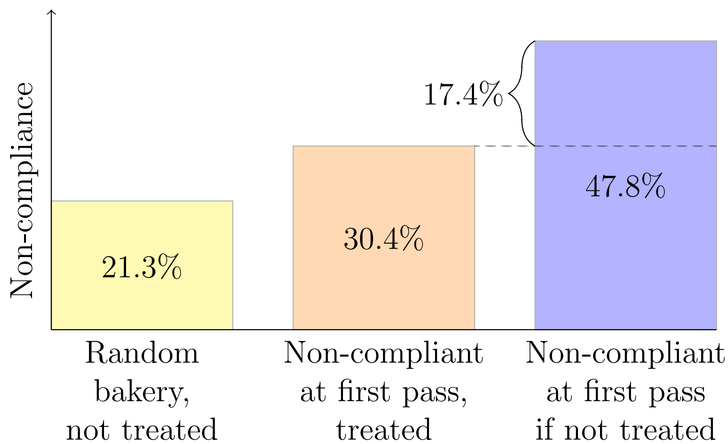
$$\widehat{ATT} = \frac{0.148 - 0.111}{0.213} = 17.4\%. \tag{4}$$

For bakeries which did not release the receipt, the treatment increases the probability of receiving the receipt in the second pass by 17.4 points. Thus, since the non-compliance rate for treated bakeries is 30.4%,[12] the non-compliance rate in absence of treatment for previously non-compliant bakeries can be estimated at 47.8%. Notice that both values are higher than the non-compliance rate we expect from a generic bakery, which is

[12]This is calculated as $\frac{\mathbb{P}\{Y_1=0,Y_2=0\}}{\mathbb{P}\{Y_1=0\}}$.

$\mathbb{P}\{Y_1 = 0\} = 21.3\%$ (see Figure 2): bakeries which are non-compliant in the first pass are *intrinsically* different in terms of compliance propensity. Compared to a random bakery, a non-compliant (and not manipulated) bakery has a probability higher by 26.5% of being non-compliant again. This provides strong evidence of illegal behavior persistence: non-compliance in the second pass is higher in bakeries which were non-compliant in the first pass, even *despite* our treatment effect.[13]

Figure 2: Non-compliance rates



In order to assess if the ATT estimated in Equation 4 is statistically significant, we run a one-sided test of $\mathcal{H}_1 : ATT > 0$ against the null hypothesis $\mathcal{H}_0 : ATT = 0$. The small sample makes asymptotic distributional assumptions unlikely: hence, rather than using McNemar's test, we observe that, assuming $\mathcal{H}_0$ holds, the number N_{01} of bakeries releasing the receipt only at the second pass is distributed according to a binomial (Sheskin, 2004):

$$N_{01} \sim \mathcal{B}(\pi, N_0)$$

where π is the *"natural"* switching rate (i.e. the probability of a random bakery changing compliance status in absence of treatment), and N_0 is the total number of bakeries not releasing the receipt at the first pass. Denoting as $\hat{\pi}$ our estimate of π, the observed significance level is therefore:

[13]Running an exact Fisher test for Table 1 allows us to reject the null hypothesis that the decisions to release the receipt at the first and the second sales are independent events ($p = 0.003$).

$$\mathbb{P}\left\{N_{01} \geq 16 | N_{01} \sim \mathcal{B}(\hat{\pi}, N_0)\right\} = 7.07\% \quad (5)$$

(see Appendix A for details). We are hence able to identify a causal effect of the treatment on the treated (ATT) with p-value $< 10\%$.

3.1. Robustness

Of the 216 sales observed in the experiment, 158 involved female vendors (who faced a female client 51,9% of the time), 58 involved male vendors (who faced a male client 55,2% the time). When the client and the vendor are of the same gender, the probability of the receipt being released drops by 13.6%, and the difference is statistically significant at the 5% level.[14] This effect is far larger than the effect of the mere gender of the client or of the vendor, which are non-significant (p-values of 30.4% and 62.21%, respectively).

In order to rule out the possibility that our main results concerning the treatment effect are driven just by the higher frequency of coincidence of genders in the first pass, we disaggregate our data as shown in Table 2.

Table 2: Data disaggregated on coincidence of genders *in the first pass.*

	Compliance status			
	N_{00}	N_{01}	N_{10}	N_{11}
Coincidence	5	11	5	33
Non-coincidence	2	5	7	40

Note: N_{ij} is a count variable for compliance i (1 = compliant, 0 = non-compliant) at period 1 and j at period 2.

In line with the effect just mentioned, we find that, in the "coincidence" case, $N_{01} > N_{10}$, while the opposite holds for the "non-coincidence" case.[15] In the absence of an effect of requests for the receipt, we would expect the magnitude of the two differences to be the same: instead, in the "coincidence" case it is three times higher than in the other one. This discrepancy is precisely what is expected in virtue of the treatment.

We compute the ATT also for different sub-samples. We find consistent results restricting the attention both to bakeries visited in the morning hours

[14]These figures are calculated using data from both passes of the experiment.

[15]Notice that the gender of the vendor in the first and in the second pass is generally unchanged (93.10% of cases): hence, if genders coincide in the first round they almost certainly differ in the second.

(54% of the sample) and to those visited in the afternoon: the ATT is always positive. Although it is higher in the morning (0.214) than in the afternoon (0.111), the difference is not statistically significant (we also observe that the rate of compliance on the first pass is homogeneous across hours of the day). We find similar results when disaggregating on the (apparent) age of the vendor in the first pass:[16] the effect of the treatment is always positive, and no significant difference in its magnitude is found. Interestingly, the effect of the treatment seems to be stronger among bakeries which were the object of the preliminary investigation (see note 11), and which were at that time non-compliant: in such a sub-sample, the estimation of the ATT is 36.36% (p-value = 0.015).[17] This seems to suggest that the request for the receipt has a stronger impact on bakeries which are *frequently* non-compliant. A plausible interpretation for this is that frequently non-compliant bakeries are precisely the ones where customers less frequently request a receipt - and hence in which one request has a stronger impact.[18] We can make a more conservative estimate of our main result, coping with the non-randomness of the experimental sample (which did not include bakeries observed as compliant during the preliminary survey) by considering the extreme assumption of a *null* response for excluded bakeries. Of all the 247 randomly selected bakeries which were involved in some phase of the study, 108 (43%) were visited for the experiment, so under this extreme assumption, the expected effect for a random bakery in Milan would be $0.43 \cdot 17.4 = 7.5\%$. We can consider this value as a lower bound to the general effect.

4. Discussion

In Section 1 we already referred to some of the channels, both psychological and purely utilitarian, through which social pressure could be affecting compliance decisions. For instance, the seller may be *ashamed* of having received a rebuke. Alternatively, he may feel *embarrassed* by the discovery that

[16] Vendors were recorded as "young" when they were attributed 30 years or less (this measure has clearly no ambition of absolute precision).

[17] **Although such bakeries on average show a lower propensity to release the receipt in the first pass and a higher propensity to react to the rebuke compared to other bakeries, these correlations are not statistically significant. Testing them through an exact Fisher test yields a p-value of 0.22 and 0.37 respectively.**

[18] We thank an anonymous referee for this intuition.

he is acting *unjustly* (and possibly, that a sense of justice is more widespread than he used to think).[19] These two different approaches can be seen as corresponding to the concepts of *collective* and *subjective* moral constraints studied in laboratory experiments by Bosco and Mittone (1997). Based on the structure of our experimental data, if the observed effect was related to *collective* moral constraints (*shame*) we would expect to find a larger impact of the rebuke when it is enacted in the presence of other clients. However, such effect cannot be disentangled empirically from possible confounding factors (e.g. it could be that in shops with more clients, sellers tend to be *ex ante* less susceptible to social pressure). Indeed, the number of clients in the shop is not significant (possibly because of the small sample size). Our results are instead consistent with the idea that the intimate feeling of injustice plays a central role, an interpretation also supported by the experimental work of Bosco and Mittone (1997).

It should be pointed out that the two agents could possibly meet two different sellers inside the shop. The available data suggests that this is not a frequent event: although among bakeries there is a large variability in the characteristics (gender, apparent age and ethnicity) of the vendors, as recorded by agents, only in 14% of cases we find a difference between the two passes. Most importantly, while in 40% of cases two or more vendors were present during the purchase, the size of the shop and the number of other clients were typically such that *any* seller would notice the request for the receipt: even in the case that our estimates are capturing a "within bakery" rather than "within seller" effect, they are what matters for policy implications (social pressure is still relevant even if felt indirectly by another seller of the shop) and presumably represent a lower bound for the "within seller" effect.

Finally, it is out of the scope of the present study to investigate specific patterns of compliance. For instance, sellers might be targeting a given level of declared sales at the end of the day. If this was the case, then our treatment should be associated with a *decreased* compliance in the following minutes, and hence we would be measuring a lower bound for the effect. Still, the experiment does not allow us to investigate the presence of more complex strategies, which are left for future research.

[19] The seller could also expect that this goes hand in hand with an increase of fiscal controls on behalf of the authority.

4.1. A social fiscal multiplier

Consider a client not receiving the receipt and asking for it. The *ATT* measures the effect of this event on the probability that, approximately 12 minutes later, another client receives a receipt. A more informative figure for policy implications would be the *number* of receipts which can be expected to be released, overall, as a consequence of that single request. The experimental data gathered allows us to make a back-of-the-envelope calculation of such number.

The number of receipts released by a bakery in the 12 minutes considered can be easily calculated, for bakeries which are compliant at the second pass, from the sequential number which is reported on each receipt:[20] in our sample, it was on average 5.5. Let η be the average number of clients making a purchase in a given bakery in the 12 minutes after the rebuke (clearly, $\eta \geq 5.5$), and recall from Section 3 that the observed rate of tax evasion 12 minutes after a request is 30.4%. Assuming for the moment that this is the rate of tax evasion of treated bakeries *during* the 12 minutes, and that correlation between compliance and the number of clients is negligible,[21] we can write

$$\eta \cdot (1 - 30.4\%) = 5.5 \qquad \text{which gives} \qquad \eta = 7.9.$$

Now, if we assume that the effect of the rebuke is constant in time, then we expect that

$$\eta \cdot ATT \approx 1.38$$

additional receipts are released in the subsequent 12 minutes.

Given the approximations involved, such estimation should be considered only as an attempt in grasping the order of magnitude of the effect. In particular, there are at least two reasons why it could be downward biased. First, assuming that the effect of the rebuke decreases with time, sales occurring *before* the 12th minute are expected to be affected by the request *even more*

[20] The presence of the sequential number on each receipt is a legal obligation. The numbers restart from 1 at the beginning of each day.

[21] We do verify that the correlation between the sequential number and the observed propensity to release the receipt is positive, even controlling for the time of the day, but this clearly does not imply a correlation of tax compliance with the number of clients.

than the sale taking place *at* the 12th minute.[22] Second, the above calculation entirely ignores the possible effect on sales occurring *after* the 12th minute.

By adding the "direct" effect of the rebuke (1 extra receipt) to 1.38, we estimate a lower bound for the *social fiscal multiplier*: at least 2.38 extra receipts are released on average when a client rebukes a seller. However, when considering a *general* effect, some further issues must be kept in mind. The effect of the treatment may be *local*: for instance, a seller who has been rebuked by a young client may, in the future, increase compliance when facing young clients only. Moreover, the agents had no ambitions of representing the average client of a bakery in terms of observable characteristics and loyalty to the shop. While it is reasonable to think that a rebuke coming from a loyal customer will presumably have an *even higher* psychological impact on the seller, there is no obvious intuition for the effect of other variables, such as age. This could be an interesting topic for future research. Finally, further research could be devoted at studying the interplay between persistence of illegal behavior and reaction to social pressure. It is also worth studying *to what extent* our findings can be generalized to countries with less widespread fiscal evasion than Italy. This crucially depends on the relative importance of the different channels, considered in Section 1, through which social pressure can influence the fiscal behavior.

5. Conclusions

We estimate, through a field experiment, the causal effect of social pressure on tax compliance of shop sellers. In our experiment, social pressure takes the form of a request for the receipt, made to bakery sellers who do not spontaneously release it (which constitute a substantial portion of our experimental subjects). Through the request, we manipulate the perception of the seller concerning the "common stand" of the Italian society towards fiscal evasion. Our results are in line with the established hypothesis according to which *"compliance cannot be explained entirely by the level of enforcement"* (Alm, 1996), but that rather it also depends on behavioral factors affecting the purchaser-seller relation. In particular, we are able to show that *direct*

[22]If this is true, then the reconstructed number of clients will be biased upwards, but it can be easily shown that the resulting estimate of the fiscal multiplier is still biased downwards.

social pressure increases by 17.4% the propensity of sellers to release the receipt in the near future, and the result is significant at the 10% level. This finding also suggests the existence of a "social fiscal multiplier": every request for a receipt causes the seller to release approximately 2.4 additional ones, which in turn can translate into increased tax revenues, since they constitute a fiscal proof of the sales having happened. We also find strong evidence of persistence in tax compliance behavior: in the subpopulation of bakeries observed once in a non-compliance state, the expected compliance is lower by 26.5% ($p < 0.01\%$).

Moreover, we find that the probability of receiving a receipt is significantly lower when a client is of the same gender than the seller (-13.6%, $p < 5\%$). Since the gender of the agent was chosen independently of any characteristics of the bakery, the effect of the coincidence of genders has a causal interpretation. The explanatory power of this interaction variable is far larger than the effect of the mere gender of the client or of the vendor, which are non-significant. This finding suggests that *illegality feeds out of complicity*, the latter being scarcer when individuals belong to different social groups (in this case, defined by gender), and provides additional evidence in favor of Torgler's point of view presented above. A word of warning is however required: the experiment involved *only one* agent of each gender. Additional evidence based on experiments involving more actors would be required to confirm that what we observe is indeed a consequence of the gender matching, rather than of individual characteristics of the agents.

The policy implications of our study consist in a strong support for awareness campaigns and other instruments aimed at influencing the behavior of sellers through *soft incentives* (positive or negative): namely, the strengthening of social norms and the diffusion of best practices.

Our experimental setting enables us to measure the *short-term* effect of social pressure. This is a unique feature among field experiments on fiscal compliance, which makes the results particularly interesting for what concerns the psychology of tax compliance decisions. However, the design could easily be extended also to the study of medium term effects: experiments conducted with more than 2 agents acting consecutively, after predetermined intervals of time, may shed some light on the persistence of the effect of social pressure, a very relevant issue for policy implications. Further research could also be devoted to measuring the sensitivity of the results to changes in the location of the experiment. While we expect the results to be quite sensitive to the city or country where the experiment is run (the effect of peer pressure will for instance depend upon the initial level of compliance), the combined results of studies coming from several towns could yield a more complete picture on the phenomenon. Other design choices worth experimenting with are the type of shop, and most importantly the characteristics of the agents.

Aknowledgements

We thank Luca Stanca, Björn Frank, Enrico Rettore and Erich Battistin for providing fruitful suggestions and support. We are grateful to participants at conferences where this work was presented and anonymous referees, for their insightful comments.

Appendix A Significance

In what follows, we provide a detailed computation for Equation (5). Assuming $\mathcal{H}_0 : ATT = 0$ holds, N_{01} is distributed according to a binomial where the number of draws is equal to the total number of treated bakeries, and therefore to the number of bakeries not releasing the receipt at the first pass (N_0), while the probability of each bakery changing compliance status by pure chance is the "natural" variability π:

$$N_{01} \sim \mathcal{B}(\pi, N_0).$$

In order to estimate

$$\pi = \frac{\mathbb{P}\{c_1 = 1, c_2 = 0\}}{\mathbb{P}\{c_1 = 0\}}$$

we use the known sample statistics:

$$\hat{\pi} = \frac{N_{10}}{N_0} = \frac{12}{23} = 0.52 \implies N_{01} \sim \mathcal{B}(0.52, 23).$$

Observing a value of $N_{01} = 16$, we can therefore calculate the probability of a type I error as:

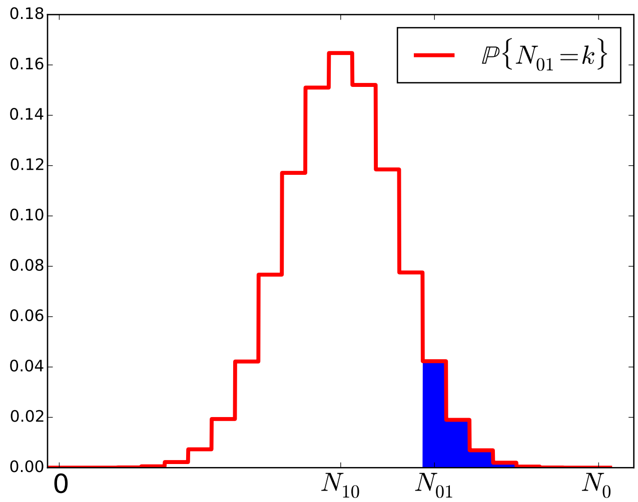
$$\mathbb{P}\left\{N_{01} \geq 16 | N_{01} \sim \mathcal{B}(\hat{\pi}, N_0)\right\}.$$

The final estimate for the p-value is hence:

$$\sum_{k=N_{01}}^{N_0} \binom{N_0}{k} \hat{\pi}^k (1-\hat{\pi})^{N_0-k} = \sum_{k=16}^{23} \binom{23}{k} 0.52^k \cdot 0.48^{23-k} = 0.0707$$

corresponding to the blue area in Figure 3.

Figure 3: Probability distribution $\mathcal{B}(\hat{\pi}, N_0)$.



References

Allingham, M., Sandmo, A., 1972. Income tax evasion: A theoretical analysis. Journal of Public Economics 1 (3-4), 323–338.

Alm, J., 1996. What is an" optimal" tax system? National Tax Journal, 117–133.

Alm, J., McClelland, G. H., Schulze, W. D., 1992. Why do people pay taxes? Journal of Public Economics 48 (1), 21–38.

Alm, J., McClelland, G. H., Schulze, W. D., 2007. Changing the social norm of tax compliance by voting. Kyklos 52 (2), 141–171.

Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. Journal of Economic Literature 36 (2), 818–860.

Asch, S. E., 1955. Opinions and social pressure. Scientific American 193 (5), 31–35.

Bordignon, M., 1993. A fairness approach to income tax evasion. Journal of Public Economics 52 (3), 345–362.

Bosco, L., Mittone, L., 1997. Tax evasion and moral constraints: some experimental evidence. Kyklos 50 (3), 297–324.

Cowell, F. A., 1991. Tax-evasion experiments: an economist's view. In: Webley (Ed.), Tax Evasion: An Experimental Approach. Cambridge University Press, pp. 123–127.

Cummings, R. G., Martinez-Vazquez, J., McKee, M., 2001. Cross cultural comparisions of tax compliance behavior. International Studies Program Working Paper Series, at AYSPS, GSU.

Cummings, R. G., Martinez-Vazquez, J., McKee, M., Torgler, B., Dec. 2006. Effects of tax morale on tax compliance: Experimental and survey evidence. Working paper series, Berkeley Olin Program in Law & Economics.

Elffers, H., Weigel, R. H., Hessing, D. J., 1987. The consequences of different strategies for measuring tax evasion behavior. Journal of Economic Psychology 8 (3), 311–337.

Erard, B., Feinstein, J. S., 1994. Honesty and evasion in the tax compliance game. The RAND Journal of Economics 5 (1), 1–19.

Erard, B., et al., 2002. Compliance measurement and workload selection with operational audit data. In: Internal Revenue Service research conference, George Washington University, Washington, DC, June. pp. 11–12.

Fabbri, M., Hemels, S., 2013. 'Do You Want a Receipt?'Combating VAT and RST Evasion with Lottery Tickets. Intertax: international tax review 41 (8/9), 430–443.

Falk, A., Fischbacher, U., 2002. "Crime" in the lab-detecting social interaction. European Economic Review 46 (4), 859–869.

Falk, A., Ichino, A., 2005. Clean evidence on peer effects. Journal of Labor Economics 24 (1), 39–57.

Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. Human Nature 13 (1), 1–25.

Fellner, G., Sausgruber, R., Traxler, C., 2013. Testing enforcement strategies in the field: Threat, moral appeal and social information. Journal of the European Economic Association 11 (3), 634–660.

Fortin, B., Lacroix, G., Villeval, M.-C., 2007. Tax evasion and social interactions. Journal of Public Economics 91 (11), 2089–2112.

Galbiati, R., Zanella, G., 2012. The tax evasion social multiplier: Evidence from italy. Journal of Public Economics 96 (5), 485–494.

Gordon, J., 1989. Individual morality and reputation costs as deterrents to tax evasion. European Economic Review 33 (4), 797–805.

Graetz, M., Wilde, L., 1985. The economics of tax compliance: fact and fantasy. National Tax Journal 38, 355–363.

Grasmick, H. G., Bursik Jr, R. J., 1990. Conscience, significant others, and rational choice: Extending the deterrence model. Law and Society Review 24 (3), 837–861.

Groenland, E. A., Van Veldhoven, G. M., 1983. Tax evasion behavior: A psychological framework. Journal of Economic Psychology 3 (2), 129–144.

Halla, M., 2012. Tax morale and compliance behavior: First evidence on a causal link. The BE Journal of Economic Analysis & Policy 12 (1).

Kidder, L. H., Bellettirie, G., Cohn, E. S., 1977. Secret ambitions and public performances: The effects of anonymity on reward allocations made by men and women. Journal of Experimental Social Psychology 13 (1), 70–80.

Kirchler, E., 2007. The economic psychology of tax behaviour. Cambridge University Press.

Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S., Saez, E., 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica 79 (3), 651–692.

Kolm, S.-C., 1973. A note on optimum tax evasion. Journal of Public Economics 2 (3), 265–270.

Levitt, S. D., Wolfram, C. D., 1997. Decomposing the Sources of Incumbency Advantage in the U.S. House. Legislative Studies Quarterly 22 (1), pp. 45–60.

Lewis, A., 2011. The social psychology of taxation. British Journal of Social Psychology 21 (2), 151–158.

Paternoster, R., Saltzman, L. E., Waldo, G. P., Chiricos, T. G., 1983. Perceived risk and social control: Do sanctions really deter? Law and Society Review 17 (3), 457–479.

Posner, E. A., 2000. Law and social norms: The case of tax compliance. Virginia Law Review 86 (8), 1781–1819.

Reis, H. T., Gruzen, J., 1976. On mediating equity, equality, and self-interest: The role of self-presentation in social exchange. Journal of Experimental Social Psychology 12 (5), 487–503.

Schwartz, R. D., Orleans, S., 1967. On legal sanctions. The University of Chicago Law Review 34 (2), 274–300.

Sheskin, D., 2004. Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC.

Singh, B., 1973. Making honesty the best policy. Journal of Public Economics 2 (3), 257–263.

Slemrod, J., 2007. Cheating ourselves: The economics of tax evasion. Journal of Economic Perspectives 21 (1), 25–48.

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

Torgler, B., 2002. Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. Journal of Economic Surveys 16 (5), 657–683.

Wärneryd, K.-E., Walerud, B., 1982. Taxes and economic behavior: Some interview data on tax evasion in Sweden. Journal of Economic Psychology 2 (3), 187–211.

Weigel, R. H., Hessing, D. J., Elffers, H., 1987. Tax evasion research: A critical appraisal and theoretical model. Journal of Economic Psychology 8 (2), 215–235.

Wenzel, M., 2001. Misperceptions of social norms about tax compliance (2): A field-experiment. Centre for Tax System Integrity Working Paper.