Regular article

Compliance and accountability-seeking: Evidence from a field experiment in Argentina

Krzysztof Krakowski [a] [ID],*, Lucas Ronconi [b,c,d] [ID],**

[a] *King's College London, United Kingdom*
[b] *University of San Martín, Argentina*
[c] *IZA, Germany*
[d] *PEP, Kenya*

ARTICLE INFO

ABSTRACT

Does compliance with low-cost civic duties increase accountability-seeking behaviors among citizenry? We address this question by conducting a field experiment at train stations in Buenos Aires. We create exogenous variation in compliance with paying the public transportation fare by: (i) highlighting sanctions for non-compliance and ii) appealing to social norms whereby 90% passengers pay the ticket. We find that both sanctions and norms treatments raise compliance. However, only appeals to social norms make treated passengers more willing to sign a petition demanding quality public transportation service—our measure of accountability-seeking behavior. To probe the mechanisms explaining these patterns, we show that compliance invoked by adherence to norms makes subjects feel more entitled to demand accountability and trust the government to a greater extent. Our findings suggest that raising compliance through appeals to social norms may thus have wider benefits for civic behaviors.

1. Introduction

A common paradox of many developing states is that they badly need revenues, yet oftentimes *intentionally* refrain from collecting taxes and choose to *not* enforce compliance (Chaudhry, 1997; Holland, 2015). One proposed answer to this puzzle is that the states buy themselves political quiescence by tolerating noncompliance and thus undermining citizens' right to seek accountability (Holland, 2016). Accountability-seeking refers to a non-electoral mechanism of control of political authorities by citizens, civic associations, movements, and the media who monitor public officials, expose bad governance, and demand exposed deficiencies or wrongdoings to be redressed (Peruzzotti and Smulovitz, 2006, 10).

A growing literature has documented that compliance breeds accountability. Citizens in developing countries have been shown to demand more say in what the government does when it tries to tax them (Paler, 2013; Weigel, 2020). Political historians have long argued that European monarchs who needed tax revenue had to cede political control in exchange for tax compliance (North and Weingast, 1989; Tilly, 1992). In rentier states, leaders provide goods to citizens for free in exchange for quiescence (Waterbury, 1997; Bablawi and Luciani, 1987).

In this study, we extend the theory of how compliance affects accountability-seeking by emphasizing that the mechanism generating compliance plays a crucial role in shaping subsequent behaviors. We hypothesize that compliance driven by coercion or the threat of sanctions is unlikely to enhance citizens' willingness to contribute individually to accountability-seeking efforts, given the inherent collective action dilemma involved in such behavior. In contrast, compliance fostered through appeals to collective responsibility, such as the voluntary water conservation efforts in Cape Town during the 2017–2018 drought (Matikinca et al., 2020), may encourage individuals to invest greater effort in holding institutions accountable.

A vast literature on fiscal contracts identifies two dominant motivations for complying with civic duties such as paying taxes or maintaining public spaces: fear of sanctions and "tax morale"—a norm

---

[1] While we refer to types of individuals, it is also possible that each person possesses both *homo-economicus* and *homo-reciprocans* tendencies (or sides). In this case, our first treatment (details forthcoming) is expected to activate the *homo-economicus* side by priming sanctions, while our second treatment, appealing to social norms, is likely to activate the *homo-reciprocans* side. Fig. 1 illustrates the implication of the distinction between human types or sides for our hypotheses, highlighting the consistency of our predictions under both assumptions.

of fairness and a related sense of duty to contribute to society (see, e.g., Hallsworth, 2014; Horodnic, 2018; Dularif and Rustiarini, 2021). Individuals fitting the *homo-economicus* type are expected to respond most strongly to the threat of sanctions, as their rational self-interest leads them to calculate that the cost of non-compliance may outweigh the benefits of evasion or free-riding. In contrast, individuals of the *homo-reciprocans* type comply due to their intrinsic sense of shared responsibility and motivation to contribute to public goods as part of a collective effort (see Fehr and Gächter, 1998).[1]

We propose that the type of compliers and their underlying motives are critical in determining whether compliance translates into subsequent accountability-seeking behavior. The "taxation-representation" model (Peruzzotti and Smulovitz, 2006; Bates and Lien, 1985) suggests that citizens contribute financially to public services more promptly if the government allows them to monitor the public goods provision (Porumbescu et al., 2017) and provides satisfactory service. Quasi experimental evidence from Brazilian municipalities shows that the government puts more effort to improve the quality of service provision when its funding predominantly relies on citizens' taxes (Gadenne, 2017). This pattern suggests that the government anticipates taxpayers' demand fo accountability.

However, the proposed increase in accountability-seeking related to compliance should only occur when compliant behavior is motivated by a shared sense of fairness and trust in broad participation (see Dickson et al., 2022; Ortega et al., 2016), as is the case for *homo reciprocans*. When individuals perceive compliance as part of a broader social contract — where contributing to public goods is both a duty and a right — they are more likely to view accountability-seeking as a natural extension of their civic responsibilities. In this context, other homo reciprocans are also expected to uphold these norms, making authorities likely to respond in turn. The combined expectations of horizontal (citizen-citizen) and vertical (citizen-state) reciprocity drive the resulting accountability-seeking dividends. In contrast, when compliance is driven by the threat of monetary penalties, contributors need not see accountability-seeking as their right or duty. Instead, it becomes a classic collective action dilemma, where individual efforts appear irrational and are thus unlikely to be observed among *homo economicus* types.

To test these hypotheses, we conduct a field experiment at metro-train stations located in the Buenos Aires metropolitan area in which we exogenously increase compliance with a low-cost civic duty: payment for train tickets (see Dai et al., 2018). To raise compliance with fare payment, we appeal to the fear of sanctions and invoke adherence to social norms, which have been documented as the most common reasons for which people comply with their basic civic duties around the world.[2] Crucially, our experimental protocol allows us to influence different types of compliers — homo economicus and homo reciprocans — *separately*. We measure the subjects' investments in accountability-seeking by eliciting their willingness to sign a petition demanding quality public transportation service. The petition requests the enforcement of the obligation that the public transportation administration has to provide a minimum of services even during strikes, as stipulated by the Argentine Law 25877, article 24. Subjects who decide to sign the petition must provide their name and surname to our confederates, which makes their actions costly.

Consistent with our theoretical intuitions, we find that appealing to the fear of sanctions and "nudging" people to adhere to civic norms increases compliance with paying the public transportation ticket. However, only appeals to norms also make passengers more willing to sign the petition demanding quality public services—an important indication of subjects' propensity to invest effort to hold their governors accountable. We show that compliers motivated by norms are more

likely to believe they have the right to demand quality public services. They also trust the government to a larger degree. We conceptualize these patterns as expressions of individuals' beliefs in vertical reciprocity (citizen-state).

Our study contributes to the literature on tax compliance and fiscal contract by, first, refining the established argument according to which compliance breeds accountability (Paler, 2013; Weigel, 2020). We show that this logic is only true for compliers motivated by normative — and not self-regarded — considerations. This finding has important implications for policy makers who design institutions that promote compliance and aim at increasing citizens' contributions to public goods. Threatening citizens with sanctions may bring greater *targeted* results in the short run.[3] Yet, appealing to social norms could have broader, long-tern dividends across various domains of civic behaviors.

Second, we subject the "compliance breeds accountability" logic to a rigorous test by examining an everyday form of compliance that entails a *direct* exchange of payment for services. This setting contrasts with much of the fiscal contract literature, which has focused on the *implicit* exchange involved in tax payments and government service delivery. The implicit — and thus delayed and partially unobservable — nature of the latter exchange plausibly intensifies citizens' urge to monitor how the government uses their tax contributions. Yet, it is less apparent whether compliant citizens would also be inclined to oversee government actions *after* they have already received the services they paid for—as is the case with public transportation. Seeking accountability in this case ensures that public transportation users, *in general*, receive better value for their money.

The remainder of this paper is structured as follows. Section 2 reviews the literature on civic compliance, highlighting the fear of sanctions and cooperative norms as two well-documented and prevalent drivers of compliant behavior. Based on this discussion, we derive our first hypothesis, which replicates an established finding in the context of our experiment. Section 3 introduces our novel theoretical framework, which connects different types of compliers to distinct patterns of accountability-seeking behavior. Section 4 details our research design, including sampling and measurement strategies as well as the experimental protocol. Section 5 presents our results, first demonstrating the successful replication of established findings and then providing empirical support for our new theoretical predictions. We further validate these findings through a series of permutation and falsification tests. Section 6 explores the mechanisms underlying our results, offering suggestive evidence consistent with the theoretical logic of our hypotheses. Finally, Section 7 concludes with a discussion of broader implications, including potential policy applications and directions for future research.

2. What explains civic compliance?

The literature on compliance has identified two broad categories of explanations of why people comply with their civic duties (for a recent review, see Mascagni 2018). These explanations relate to the role of (i) sanctions, and (ii) normative concerns.

The first explanation proposes that people comply with civic duties if their non-compliance could be easily detected and punished by some legal authorities (Alm et al., 2012; Alm, 2012; Kirchler et al., 2008; Ritsatos; Torgler and Werner, 2005; Katz and Owen, 2013; Riahi-Belkaoui, 2004). This mechanism aligns with another robust finding from the fiscal contract literature, which establishes that compliance increases with administrative capacity (Nurkholis et al., 2020; Andriani, 2016;

---

[2] See, e.g., Alm et al. (2017), Hallsworth (2014), Bursztyn et al. (2019) and Saulitis (2023).

[3] The sanctions appeals produce 12.1 percentage-point increase in ticket payment vis-á-vis 5.4 percentage-point increase related to the norms appeal. If sanctions and norms-based approaches are implemented similarly in real-life settings, the resulting changes in compliance levels may reflect these differences in aggregate effectiveness.

Chan et al., 2018; Daude et al., 2012; Ortega et al., 2016; Torgler, 2003; Levi and Stoker, 2000; Kondelaji et al., 2016; Ibrahim et al., 2015; Leonardo, 2011).[4]

The second explanation posits that people comply with civic duties because they are influenced by social and moral norms of their communities that stigmatize non-compliance and other forms of uncivic behaviors (Andreoni et al., 1998; Falk and Fischbacher, 2006; Fergusson et al., 2019; Alm et al., 2017; Ritsatos; Kirchler et al., 2010; Frey and Torgler, 2007; Pruckner and Sausgruber, 2013).

A series of field experiments has tested the effectiveness of low-cost "nudging" interventions that authorities could use to increase civic compliance through the above channels. Castro and Scartascini (2015), for example, find that authorities can increase tax compliance by 5 percentage points by simply emphasizing possible fines related to tax evasion. Hallsworth et al. (2017), in turn, find that appeals to social norms of compliance can allow tax authorities to improve overdue tax collection by the margin of 2–5 percentage points. Regarding the relative effectiveness of sanctions vis-á-vis social norms approaches, the former has generally been found to work better, improving compliance by larger margins (see Hallsworth, 2014; Horodnic, 2018; Dularif and Rustiarini, 2021). Building on this evidence, we formulate our first hypothesis, where we expect to replicate the well-established finding:

H1: The fear of sanctions and appeals to civic norms will increase compliance with low-cost civic duties.

3. Compliance and accountability

Compliant citizens plausibly expect tangible returns, particularly in the form of high-quality public services. These expectations build on the notion of vertical reciprocity between the citizens and the state. When the expectations are met, citizens are more likely to adhere to the rules of good citizenship (Armand et al., 2021). Conversely, a series of lab experiments showed that when participants learned that tax revenues would be destroyed by a central authority, their compliance dropped sharply (Andrighetto et al., 2016; Steinmo and D'Attoma, 2021).

The causal relationship between compliance and accountability may operate in both directions. According to the "taxation produces representation" argument, compliance itself can strengthen demands for greater oversight, pressing governors to be more responsive to citizens' needs and ultimately increasing accountability (Ross, 2004; Peruzzotti and Smulovitz, 2006; Paler, 2013; Weigel, 2020). While the precise mechanism remains debated (Gadenne, 2017), tentative evidence suggests that fairness concerns and reciprocity expectations may drive these changes (Ronconi, 2019).

We propose to qualify the latter relationship by arguing that the accountability-seeking dividends of compliance vary with the underlying motivation which prompt people to comply with civic duties. Our theory builds on an observation that different *types* of individuals may respond to different types of appeals to comply with civic duties. Following Fehr and Gächter (1998), we distinguish between (i) *homo-economicus* and (ii) *homo-reciprocans* types.

Homo economicus engage in cooperative behaviors, including compliance with civic duties, based on rational calculations aimed at maximizing their payoffs. They weight the expected costs of non-compliance due to fines against the gains from evasion. By contrast, homo reciprocans are naturally inclined to reciprocate cooperative behaviors, often beyond immediate self-interest. They tend to exhibit prosocial behavior when they observe others doing the same. This behavior is reinforced by such mechanisms as altruistic punishment, where homo

reciprocans incur a cost to punish those who violate cooperative norms, thus sustaining cooperation within a group (Fehr and Gächter, 2002).

Building on the distinction between homo economicus and homo reciprocans, we propose that the two main strategies for promoting compliance — fear of sanctions and appeals to cooperative norms — target these types to different degrees. The sanctions-based approach relies on the threat of audits and fines, increasing the expected monetary costs of non-compliance—thus primarily influencing homo economicus. In contrast, the norms-based approach relies on peer pressure to encourage conformity with majority behavior, making it more effective for homo reciprocans. In our context, highlighting that 90% of public transportation users comply with fare payment — compared to the ~30% baseline observed at the studied train stations — helps update individuals' beliefs about the prevalence of cooperators, reinforcing trust among conditional cooperators (Bicchieri, 2005). By highlighting prosocial behavior as the norm, these messages also evoke fairness considerations, implicitly framing compliance as the right thing to do, particularly when others also participate.[5]

Hypotheses. Why should the sanctions-based and norms-based approaches to compliance have different effects on individual investments in accountability-seeking? Sanctions enforce compliance through external penalties, appealing primarily to homo economicus, who make cost-benefit calculations. However, accountability-seeking is a collective action problem where free-riding is pervasive, making individual participation irrational unless the personal benefits outweigh the costs. Homo economicus will engage only when the expected cost of inaction exceeds the cost of participation, meaning that compliance driven by sanctions should not translate into a greater willingness to pursue costly prosocial actions like accountability-seeking.

While homo economicus may value accountability as a mechanism to ensure better public services, this preference is independent of their compliance behavior. Compliance does not affect the material costs and benefits of personally contributing to accountability-seeking efforts. Since individual participation in accountability-seeking is costly and is rarely associated with selective incentives, homo economicus types are unlikely to engage.

We thus expect that:

H2: Compliance with low-cost civic duties invoked by the fear of sanctions will *not* increase subsequent participation in collective action to monitor the quality of government service provision.

The norms-based approach fosters compliance by informing individuals about others' adherence, appealing to homo reciprocans, who value fairness and reciprocity. Unlike homo economicus, who prioritizes self-interest and free-riding, homo reciprocans perceive public goods as reciprocal arrangements—not only among citizens who use them (if fellow passengers pay for the metro, so should I, and vice versa), but also between the state and its citizens. Homo reciprocans expect fairness in both directions: just as individuals feel obligated to contribute, they also believe that public institutions must uphold their duty to provide quality services in return. The resultant sense of entitlement to demand accountability plausibly motivates individual investments in contributing to this outcome.

Importantly, the state's responsiveness to these demands depends on broad participation in accountability-seeking. Thus, homo reciprocans may see it not only as a right of those who comply but also as a civic duty to ensure that public authorities uphold their obligations

---

[4] A number of studies has tested whether altering beliefs about the capacity of the state affects compliant behaviors, confirming the positive link (e.g Coleman, 1996; Slemrod et al., 2001; Dickson et al., 2022; Gonzalez-Navarro and Quintana-Domeque, 2013).

[5] The fear of monetary fines can also influence homo reciprocans, just as peer pressure and informal sanctions (such as scorn or ostracism) may impose economic costs that factor into homo economicus' calculations. However, we argue that fines have a *stronger* impact on homo economicus than on homo reciprocans, while social norms exert a *greater* influence on homo reciprocans than on homo economicus (see Elster, 1999; Fehr and Gächter, 1998).

toward all contributors. Crucially, norms-based interventions reinforce individuals' belief in collective efficacy, reassuring homo reciprocans that others will also engage—thus reinforcing their own willingness to participate. Adida et al. (2020) illustrate this mechanism in a field experiment in Benin, demonstrating that voters hold the government accountable only when they know that performance information is widely disseminated, ensuring that all homo reciprocans are motivated to purse costly accountability-seeking actions.[6]

This discussion leads us to our final hypothesis, which proposes:

H3: Compliance with low-cost civic duties motivated by appeals to civic norms will increase subsequent participation in collective action to monitor the quality of government service provision.

The hypotheses we presented assume that human beings can be categorized into two types, as either homo economicus (HE) or as homo reciprocans (HR).[7] The work of Fehr and Gächter (2000a,b), Dohmen et al. (2009), and Murphy et al. (2011) could be taken as supporting this categorization into types of human beings (i.e., HE or HR).[8] An alternative assumption is that every human being has two sides, an HE and an HR side, and only one side is active at a time depending on the stimulus received. The work of Hodgson (2012) and Lindenberg et al. (2021) supports the view that human beings have an HE and an HR side.

What are the implications of this distinction for our proposed theory? In Fig. 1, we present our hypotheses under both assumptions regarding human cooperative dispositions, as the empirical question of (non)cooperative types or sides remains unresolved. Fig. 1 outlines predictions based on sanction- and norm-based approaches and their effects on compliant behavior and subsequent accountability-seeking. These predictions are framed in terms of our operational definitions: metro ticket payment as a measure of compliance and petition signing for quality public transportation as a measure of individual accountability-seeking (details momentarily). Crucially, our key expectation remains the same under both assumptions: norm-induced compliance should lead to greater accountability-seeking, whereas sanction-induced compliance should not—irrespective of whether HE and HR represent different types of human beings or different sides of the same individuals.

4. Design

4.1. Setting

We conducted a field experiment at metro-train stations in the Buenos Aires Metropolitan Area between October 4 and December 17, 2021. We exposed metro commuters traveling towards the periphery of the city to messages meant to induce them to pay the metro ticket. To make readership of the message high, we placed a research assistant at the entry of the train station wearing a T-shirt with the treatment message and handing out flyers to passengers with the same message (see the lower panel of Fig. 2 below).

Fig. 2 represents a typical train station and shows where each of the research assistants (RA) was located. RA1 was located at the entrance,

few meters before passengers decide whether to pay or dodge the fare by entering the platform through the "emergency" door. RA1 delivered the treatment message combining a flyer and T-shirt. Figure A2 in the Appendix shows some real train stations. RA2 counted how many people entered the platform through the turnstile (pay fare) and how many people entered through the "emergency" door (dodge the fare). RAs 3 and 4 were on the platform and conducted a follow up survey with the population of people who entered the platform.

Throughout the fieldwork, we varied the gender of RAs in specific roles. RAs 3 and 4 interviewed every third passenger entering the platform except for people who looked unambiguously older than 65 years, younger than 16 years, handicapped or wear a police uniform. We excluded these groups because people younger than 16 cannot sign petitions and may be exempt from fare payment (e.g. school children). Likewise, most people older than 65 are pensioners who receive the minimum benefit and do not have to pay the transportation fare. Handicapped individuals and police personnel on duty are also exempt.

4.2. Types of treatment

Our goal was to increase compliance through different channels (fear of sanctions vs. normative concerns) by using two treatment messages which remind passengers that: (1) there is a fine in case of evading the ticket (*Evite Multas*); and that (2) 90% of the passengers pay the ticket (*90% de los Pasajeros Pagan Boleto*). The latter message relied on an estimation of the share of *total* passengers of public transportation that pay their tickets in the Buenos Aires Metropolitan Area. Therefore, the estimate includes subway and bus passengers (where fare dodging is almost impossible), and metro-train passengers traveling to Buenos Aires downtown, where it is also very difficult to leave the terminal station without paying the fare. The information provided in the norms treatment message was therefore truthful, and should be perceived as such by passengers whom we expect to be aware of the fact that the levels of (non-)compliance observed at one station are not representative of the overall levels of fare dodging in the entire transportation network.

Our two-treatment design builds on the idea of estimating causal effects with multiple instruments (see Mogstad et al., 2021, 2024). In our case, this setup allows us to induce exogenous changes in compliant behaviors separately for different sub-populations of interest: individuals responding to economic incentives in matters of compliance (homo economicus), and those who respond to social cues and the sense of civic duties (homo reciprocans). The control group was exposed to the presence of a research assistant without any message. Each treatment (RA with T-shirt + flyer) lasted for one hour and fifteen minutes, and then it was replaced by another treatment.

We assigned treatments to station-time units on a rotation basis, using different sequences on different dates. We developed the following schedule. On the first day, we began the morning shift (9:00 AM to 1:00 PM) with T1, then T2, and finally C; and repeated the same sequence during the second shift (2:00 PM to 6:00 PM). The sequence for the second and subsequent days were altered, as illustrated in Table 1.

4.3. Sample

There are more than 150 stations along the seven train lines in the Buenos Aires Metropolitan Area, as shown in Figure A3 in the Appendix. We randomly selected train stations, and assign them to each date, from the population of stations that ex-ante met the conditions to conduct the experiment. Each week we covered a different train line. Our original plan was to visit one station per day during 11 weeks (i.e., 55 train stations). But in some cases, a revenue protection officer arrived at the train station while we were conducting the experiment, closed the emergency door, and actively enforced ticket payment. Under these circumstances, passengers had no option to dodge the fare; the conditions to conduct the experiment were not fulfilled and

---

[6] Conversely, highlighting sanctions for non-compliance — the core of the alternative approach — may signal to homo reciprocans that others are not complying, thereby reducing the perceived utility of accountability-seeking, whose value increases with the expectation of others' engagement (Sliwka, 2007).

[7] An extension of this logic entails continuous classification, such as in Szekely et al. (2021), for example.

[8] Nagel (1995) proposed an alternative interpretation of cooperative behavior, arguing that cooperation is based on a lack of strategic reasoning and selfishness is simply strategic reasoning. She used the Guessing Game as a way of identification how deeply people reason in terms of actions and effects.
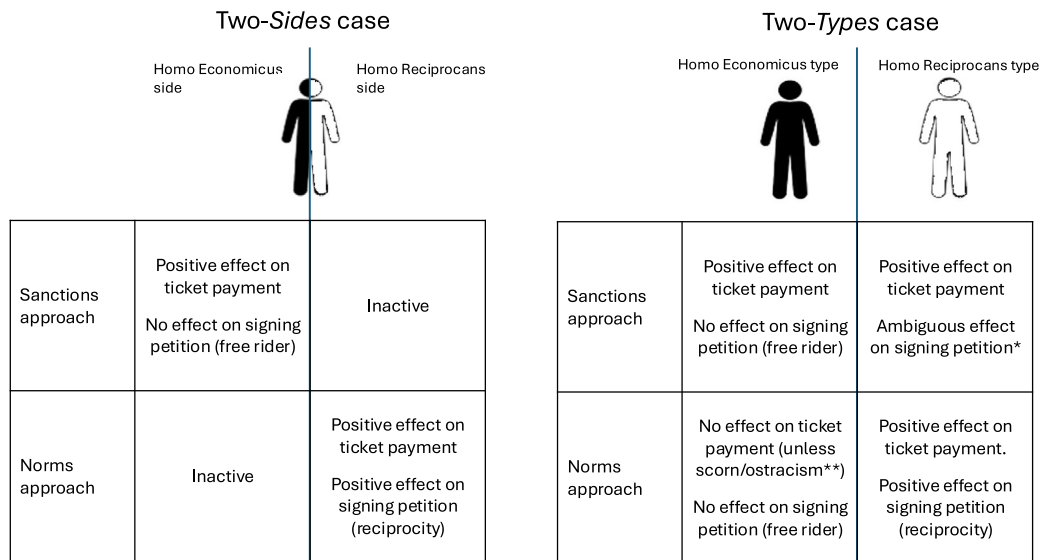
Fig. 1. Hypotheses under two assumptions about human (non)cooperative nature.
Notes: *The expected effect on signing the petition is ambiguous since paying ticket makes HR feel entitled to demand quality service, but ultimately unsure of the broader participation in accountability-seeking. Emphasizing fines may signal to HR that others are not complying and are thus unlikely to contribute to accountability-seeking. **Scorn or ostracism may impose economic costs that factor into homo economicus' calculations.

Table 1
Treatment schedule during the first week of intervention.

Schedule	Monday Station 1	Tuesday Station 2	Wednesday Station 3	Thursday Station 4	Friday Station 5
9:00 – 10:15 AM	T1	T2	C	T1	T2
10:20 – 11:35 AM	T2	C	T1	T2	C
11:40 – 12:55 AM	C	T1	T2	C	T1
Lunch					
2:00 – 3:15 PM	T1	T2	C	T1	T2
3:20 – 4:35 PM	T2	C	T1	T2	C
4:40 – 5:55 PM	C	T1	T2	C	T1

Table 2
Descriptive statistics.

Group	Observations	Paid ticket	Age	Female	Female interviewer
Control	2579	0.315	35.879	0.521	0.772
T1 (Sanctions)	2525	0.437	36.337	0.531	0.791
T2 (Norms)	2523	0.369	35.794	0.543	0.770
Diff C - T		−0.088***	−0.188	−0.0167	−0.008
Std. Err.		(0.012)	(0.297)	(0.012)	(0.010)

Notes: Significance level shown below *p<0.10, ** p<0.05, ***p<0.01.

we thus immediately ended the experiment and moved to a different train station.[9] In sum, we collected data in 62 different train stations (Appendix A.1 provides a full list and a map).

4.4. Outcomes

Inside the station, a research assistant checked whether the entering passengers paid the ticket (RA2), and two additional researchers (RA3 and RA4) asked passengers to complete a very short survey while waiting in the platform for the arrival of the train. The survey included questions about age, sex, and, crucially, whether the passenger was willing to sign a petition demanding the enforcement of a law which stipulates that a minimum of public transportation service should be provided even during strikes (for the exact question wording, see Appendix A.3). Signing the petition is our behavioral measure of respondents' costly accountability-seeking behavior. RA3 and RA4 also recorded whether a given interviewee paid the fare. We compared these compliance estimates based on the individual-level data with aggregate-level counts collected by RA2 (see Appendix A.4).

In the course of the fieldwork, some respondents refused to answer our survey and we were thus unable to measure their outcomes (beyond independently observable compliance with fare payment and basic demographics). We address this issue in three steps. First, we investigate whether response rates differ by treatment categories, detecting evidence of differential attrition (Figure A4). Second, we address the problem of differential attrition by estimating a series of selection
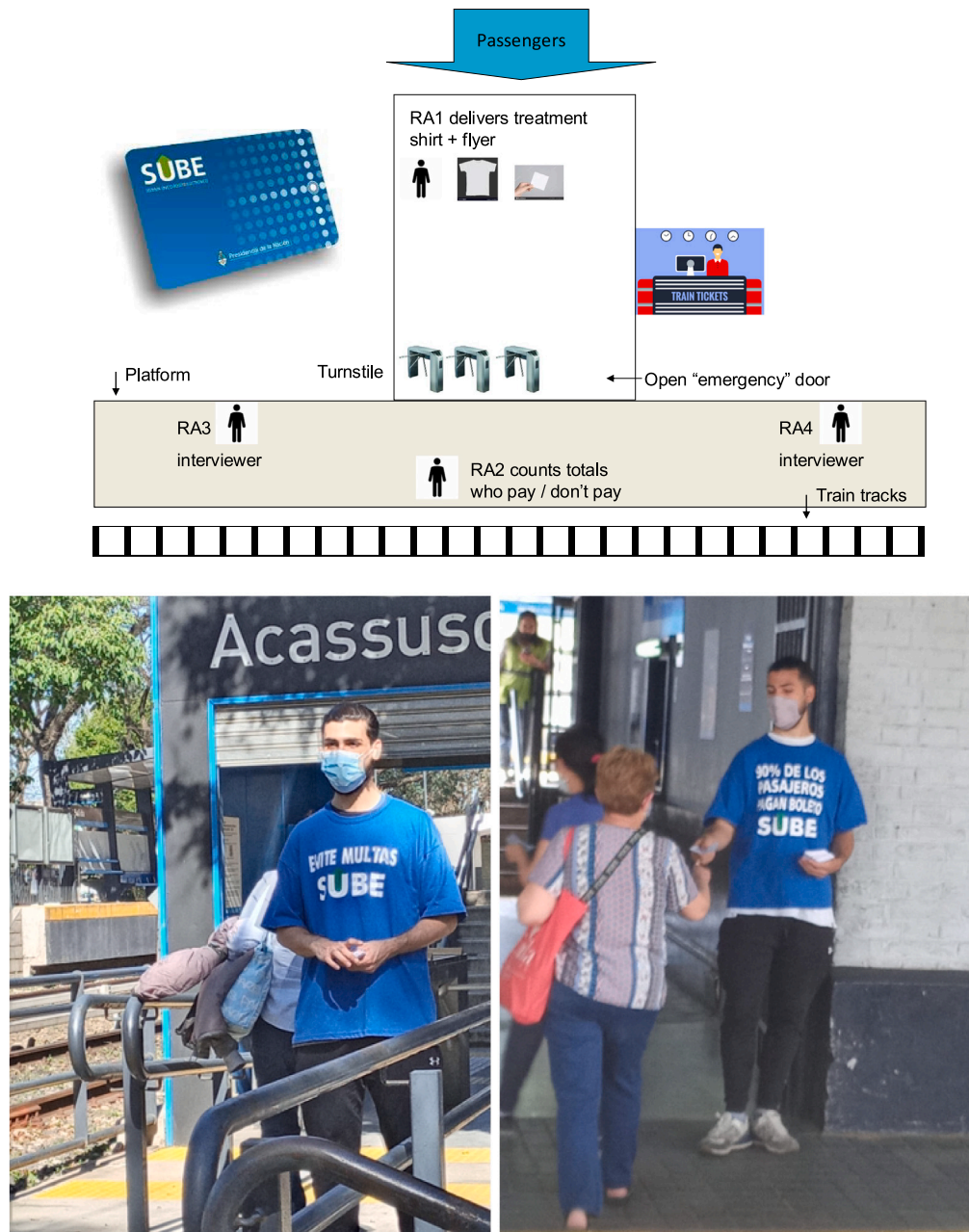
---

[9] We acknowledge that visits from revenue protection officers are not random. However, they should not be systematically correlated with our treatment. Each station was exposed to every experimental condition twice per day, with the order of treatments determined randomly within our rotation schedule. While revenue protection officers may visit certain stations more frequently or conduct inspections at specific times of the day, these (overall infrequent) interruptions to our fieldwork would still be random with respect to the treatment being implemented at the time.

Fig. 2. Experimental set-up.

Notes: The upper panel shows the experimental set-up, outlining the position of our research assistants—those delivering the treatment and those collecting compliance and survey data. The lower panel shows one of our research assistant (RA1) while he delivers Treatment 1 that appeals to the fear of sanctions (on the left) and Treatment 2 that appeals to norms (on the right).

models, following Heckman (1976) and related models with endogenous treatment and sample selection (details momentarily). Third, we impute extreme values on our petition variable for the non-respondents and examine how sensitive our estimates are to these imputations (Figure A5). Reassuringly, we find consistent results across these tests, as discussed in detail below.

5. Results

We divide the discussion of our empirical results in two parts. First, we analyze how the propensity to pay the fare varies by treatment groups. Second, we exploit the exogenous change in the propensity to pay the fare to study its effects on accountability-seeking behaviors. In all the forthcoming analyses, we restrict our sample to people 16

to 65 years of age (N=7,627). We do so because people younger than 16 cannot sign petitions and most people older than 65 do not have to pay the transportation fare. Yet, our results are robust to including these people (N=168) in the analytical sample.

5.1. Effects of treatment on compliance

Fig. 3 (left panel) compares the means of fare payment by treatment conditions. Only 31.5% of passengers in the control group paid their tickets. During Treatment 1 ("avoid fines;" the sanctions message), the share paying was 43.7%; and during Treatment 2 ("90% passengers pay their tickets;" the norms message), the share paying was 36.9%. Differences between these treatment groups and the control group are

Table 3
Treatment effects on ticket payment.

	(1) Paid ticket	(2) Paid ticket	(3) Paid ticket	(4) Paid ticket
T1 (Sanctions)	0.122***	0.111***	0.108***	0.108***
	(0.013)	(0.013)	(0.013)	(0.013)
T2 (Norms)	0.054***	0.057***	0.057***	0.056***
	(0.013)	(0.013)	(0.013)	(0.013)
Station and schedule FE	No	Yes	Yes	Yes
Passenger controls	No	No	Yes	Yes
Sample	16–65 yrs old	16–65 yrs old	16–65 yrs old	Full
N	7627	7627	7627	7795

Notes: Robust standard errors in parenthesis. Significance level shown below *p<0.10, ** p<0.05, ***p<0.01.
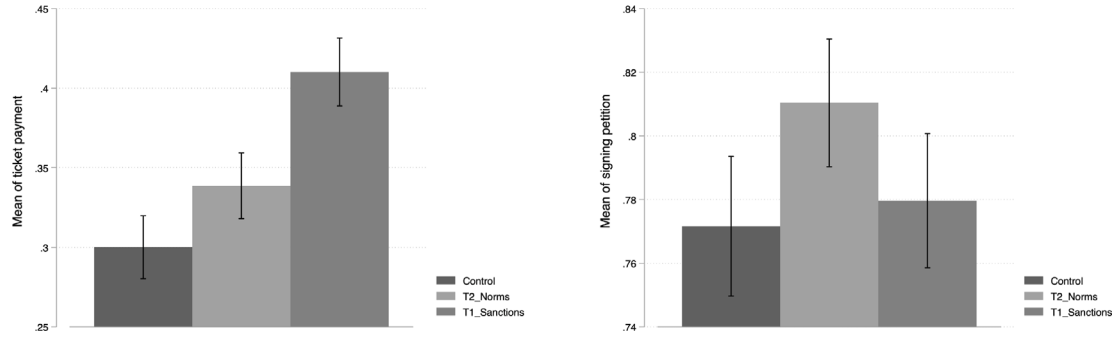

Fig. 3. Treatment, compliance, and demand for social accountability.
Notes: The figure shows the means and the accompanying 95 confidence intervals of the indicated outcomes by the treatment assignment status.

statistically significant at the 1% level (see Table 2).[10] In sum, the descriptive analysis suggests that our interventions raised compliance, as expected.

These patterns are further confirmed in the regression framework. We investigate the effects of treatment on fare dodging by estimating the following linear model:

$$P_{ijs} = \alpha_j + \gamma_s + \beta_1 T1_{js} + \beta_2 T2_{js} + \delta X_i + \epsilon_{ijs} \qquad (1)$$

where P is an indicator for paying the train fare of an individual i in a station j during a schedule s; $T1$ indicates whether the passenger entered the station when there was a sanctions message, and $T2$ indicates whether the passenger entered the station when there was a social norms message. We evaluate the effects of these treatments with reference to the control group. X is a vector that includes the age, age squared and sex of the passenger[11]; and α_j and γ_s are train station and schedule fixed effects. We use robust standard errors.

Table 3 presents the results of these analyses. In column (1) we do not include any controls, in column (2) we add station and schedule fixed effects, and in column (3) we add the vector of passenger characteristics. Finally, in column (4) we include passengers below 16 and above 65 years of age. The results of all the specifications confirm that both sanctions and social norms interventions make people comply with low-cost civic duties at higher rates. The results are substantially unchanged if we use a probit model, instead of the linear regression.

---

[10] The similar number of individuals treated by T1 and T2 results from our randomization procedure, which assigned treatments within stations across different time windows. This approach ensured that each station was exposed to all treatments twice at randomly selected intervals within a given day. Consequently, our design accounted for potential between-station differences in commuter population size as well as time-based variations in traffic. Given this rotation, we expected the treatment groups to be closely balanced in size.

[11] We use these controls to increase the precision of our estimates (see Gelman et al., 2021). Table 2 shows that the covariates are balance with respect to the treatment assignment, including the age and sex of the passenger, and the sex of the interviewer.

5.2. Effects of (exogenous) compliance on accountability-seeking

In the next step, we examine the effects of treatment on our measure of individual contribution to accountability-seeking: signing the petition ($SignPetition$, the mean of 0.787). Fig. 3 (right panel) shows that passengers exposed to our norms treatment are more willing to sign the petition compared to the control condition. In line with our expectations, the sanctions treatment does not have the same effect.

We confirm these descriptive patterns by estimating a two-least-square (2SLS) instrumental variable (IV) regression:

$$\text{First Stage:} \quad P_{ijs} = \alpha_j + \gamma_s + \lambda T1/T2_{js} + \delta X_i + \eta_{ijs} \qquad (2)$$

$$\text{Second Stage:} \quad S_{ijs} = \alpha_j + \gamma_s + \beta \hat{P}_{js} + \delta X_i + \epsilon_{ijs} \qquad (3)$$

where S is an indicator for signing the petition by an individual i in a station j during a schedule s. We use the two different treatments as instruments of compliance in separate IV models to distinguish between two sub-populations of passengers (Mogstad et al., 2021, 2024). The models control for age, age squared, sex, sex of the interviewer, and include train station and schedule fixed effects. In addition, we estimate a reduced form OLS model in which we regress the petition outcome directly on the exogenous treatment assignment (and covariates):

$$S_{ijs} = \alpha_j + \gamma_s + \beta_1 T1_{js} + \beta_2 T2_{js} + \delta X_i + \epsilon_{ijs} \qquad (4)$$

Columns 1–3 of Table 4 present the results of these analyses. Paying the train ticket has a positive and significant effect on signing the petition when we instrument compliance with the treatment appealing to norms (column 3). It does not have the same effect when we instrument compliance with the treatment appealing to the fear of sanctions (column 2). These results are confirmed in the reduced-form model (column 1).[12]

---

[12] In Table A3, we analyze the differences between compliers in the two treatments (T1 Sanctions and T2 Norms). A potential concern is that the effect on petition signing may not be purely driven by the treatment but rather by

Table 4
Treatment, compliance, and signing the petition.

	(1) Reduced form	(2) IV = T1 Sanctions	(3) IV = T2 Norms	(4) Reduced form (Heckman)	(5) IV = T1 Sanctions (-eprobit-)	(6) IV = T2 Norms (-eprobit-)
T1 (Sanctions)	0.009			0.080		
	(0.016)			(0.054)		
T2 (Norms)	0.037**			0.168***		
	(0.016)			(0.047)		
Paid ticket		0.099	0.600**		0.365*	0.053***
		(0.129)	(0.284)		(0.188)	(0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4360	2894	2876	6126	4106	4083

Notes: Robust standard errors in parenthesis. Significance level shown below *p<0.10, ** p<0.05, ***p<0.01.

An alternative design to test our claim is to consider T1 (Sanctions) as the control condition, given that it represents compliance without appeals for collective behavior. T2 (Norms) then serves as the treatment condition, capturing compliance with such appeals. This comparison is descriptively illustrated in Fig. 3 (right panel), where one can visually compare bars 2 and 3. We further confirm these patterns in a regression framework in Table A2 in the Appendix. We consistently find that compliance with appeals for collective behavior is associated with an increased likelihood of signing the petition compared to compliance without such appeals.

5.3. Accounting for differential attrition

In the course of our fieldwork, we approached 7,795 passengers with requests to answer our survey. 27.4% of the targets refused to take part in the survey. As a result, for part of the sample, we were unable to measure the main outcome of interest: willingness to sign the petition.

Below, we assess whether the non-response rate varied by treatment assignment. In Figure A4, we compare response rates across experimental conditions. There is clear evidence of differential response rates with subjects exposed to any of our treatment messages being more likely to answer the survey. This is unsurprising, given that people in the control condition were less likely to pay their fare and, as a consequence, could have felt uneasy to talk to strangers (e.g. if they thought those could reproach their fare evasion).[13]

Two different approaches can be taken here. The first approach is to consider non-participation in the survey as a differential attrition problem; and hence, to focus on willingness to sign the petition as the only dependent variable. This is the approach we follow below. A second approach is to note that participating in a survey about public transportation service conducted by members of a public university is a proxy for civic engagement and accountability-seeking behavior. Differential response rates in the latter case is thus consistent with the argument we put forward in the study. Paying the metro fare makes people more likely to civically engage.

Following the first, more conservative approach, we address the differential attrition problem in two ways. First, we estimate a Heckman selection model (Heckman, 1976). The model consists of two

equations. The first equation predicts the selection into answering our survey, and thus the observability of the outcome. The second equation regresses the outcome on the covariates of interest. Note that the controls used in Table 4 (columns 1–3) and the treatment assignment, which reduces non-response, as shown above, are included among the covariates in both equations.[14] Importantly, the Heckman model recognizes that unobserved factors (e.g. fearfulness or prosociality) may affect both the outcome and the probability of selection in the sample, thus introducing bias to the estimates of interest. These unobserved factors are contained in the residuals of both equations. According to Heckman, this bias can be corrected in two steps: first, by computing the expected value of the error term from the first equation conditional on the covariates predicting selection in the sample, and, second, by including this term in the main empirical model. We implement this correction in column 4 of Table 4. The table presents the result from the second equation of the Heckman model. The model reports reduced-form estimates.

In a similar spirit, we estimate a series of probit regressions that include an endogenous treatment (fare payment) and account for the fact that the data are subject to endogenous sample selection. In these models, we can instrument the endogenous treatment with exposure to our experimental conditions. Again, the models allow us to address the potential problem that unobserved factors that influence the choice of fare payment may be correlated with the unobserved factors that affect the choice of answering our survey. The models include the same control variables that we use in our main regressions (columns 1–3 of Table 4). The results are consistent with our main findings (column 5–6 of Table 4). The estimates refer to the average treatment effect on the treated (ATET).

Second, we address the problem of the missing data in the outcome variable by imputing extreme values for respondents who refused to answer our survey. We independently observed these individuals' assignment to treatment as well as their compliance with fare payment. As explained above, our research assistants also approximated these subjects' basic demographic characteristics. To evaluate how sensitive our results are to the loss of non-respondents, we first assume that they were all willing to sign the petition. Such an imputation is likely to underestimate the effect of our treatment. In the second step, we assume that none of the non-respondents was willing to sign the petition. This imputation, by contrast, plausibly overestimates the effect of treatment. Figure A5 shows the variation in estimated treatment effects for the observed outcome and the imputed ones (both for upper and lower-bound imputations). The effect on the norms treatment is marginally statistically insignificant in the case of upper-bound imputations; yet, the size and direction of the effect is in line with the previous results.

---

compositional differences among compliers. We partially address this issue by including controls in our regressions and exploring contextual heterogeneities (details in Section 6). Additionally, to further mitigate this concern, Table A3 demonstrates that observable characteristics do not significantly differ between the two groups of compliers. In other words, compliers in T1 and T2 are statistically indistinguishable in terms of gender, age, and the socioeconomic characteristics of the neighborhoods where they boarded the train, as measured by nighttime luminosity and population density.

[13] We took a number of steps to ensure that passengers would not make connections between the RA delivering the treatment and the RAs interviewing people on the platform. Most importantly, the RA at the entrance to the station was dressed in the transportation company uniform, while the others RAs were dressed casually.

[14] Our research assistants estimated gender and age also for subjects who refused to take the survey. Naturally, the estimates of age are more prone to error that the estimates of gender. However, the results are unchanged if we exclude the age variable from the selection models.

Table 5
Treatment, compliance, and signing the petition about animal rights (falsification test)

	(1) Reduced form	(2) IV = T1 Sanctions	(3) IV = T2 Norms
T1 (Sanctions)	0.024 (0.033)		
T2 (Norms)	0.005 (0.033)		
Paid ticket		0.200 (0.353)	0.011 (0.283)
Controls	Yes	Yes	Yes
N	1146	748	772

Notes: Robust standard errors in parenthesis. Significance level shown below *p<0.10, ** p<0.05, ***p<0.01.

5.4. Falsification tests

To use our treatment messages as an instrumental variable for compliance, we must invoke five assumptions. First, we must assume that our treatment is a relevant instrument of compliance, the pattern we documented in Table 3 (first-stage assumption). Second, we must assume that people comply with treatment assignment, that is, they cannot choose to be treated if assigned to a control condition, and vice versa (monotonicity assumption). Third, we must assume that there are no spillover effects of our treatment messages onto untreated individuals, e.g. those who traveled on later times of the day in the previously treated station (stable unit treatment value assumption (SUTVA) assumption). Fourth, we must assume the treatment assignment and our outcomes do not have some common unobserved causes (exogeneity assumption). Assumptions (2) to (4) are easily justifiable in our case, given the random assignment to treatment groups. Fifth, and more problematic, we must assume that the treatment affects accountability-seeking through its effect on compliance — paying the ticket — rather than through other channel(s) (excludability assumption).

While we cannot directly test the excludability assumption, we propose two falsification tests that help us rule out possible violations of the exclusion restriction assumption. A key concern is that the treatment messages could have a *direct* effect on signing the petition, independent of paying the metro ticket.

Animal rights' petition. Our first falsification tests exploits the fact that during the last three weeks of our fieldwork we modified the second part of the experiment. Instead of asking passengers whether they would like to sign a petition demanding quality public transportation service, we asked whether they would like to sign a petition demanding the prohibition of using animals for experimental purposes (see Appendix A.3).

According to the accountability logic outlined above, ticket payment should be unrelated to signing the petition requesting a ban on animal testing if the excludability assumption is met. Simply put, the ban on animal testing petition is unrelated to the quality of public services. If, however, the excludability assumption is violated, we could observe an effect of treatment (via ticket payment) on signing the animal testing petition, for instance, due to the fact that our treatment raises expectations of collective efficacy more broadly.[15]

Table 5 replicates the models from Table 4 (columns 1–3), but using data from the final three weeks of the fieldwork. The dependent variable is an indicator equal to 1 if the person signs the petition to ban animal testing, and 0 if they explicitly choose not to sign ($SignPetitionFT$, the mean of 0.777). The table shows no effect on signing the animal testing petition. This result builds confidence in the validity of the excludability assumption.

Survey experiment. Our second falsification test makes use of a new sample of university students from the University of Buenos Aires. These student were exposed to our treatment messages during an online survey. We incentivized participation in this survey by offering the students a chance to win a T-shirt in the lottery. The T-shirts were the same as those we used in the main experiment: they contained (i) the sanctions message, (ii) the norms message, and (iii) no message at all (see Fig. 4). We randomly showed one of these T-shirts to survey participants as a preview of their potential prize. After showing these T-shirts, we asked respondents whether they were willing to sign the petition demanding the enforcement of the law which stipulates that a minimum of public transportation service should be provided even during strikes. We used the same wording of the train-station questionnaire.

The crucial difference between the online and station-based experiments is that the online treatments cannot affect *actual* compliant behavior: there is no fare to pay. Therefore, if our treatments have any effect on propensity to sign the petition online, this would indicate that the exclusion restriction assumption is likely to be violated.

To estimate the necessary sample size for our online falsification test, we conducted power calculations using effect size and standard deviation inputs based on field experimental data, ensuring realistic and context-specific estimates. We focused on two-group comparisons to isolate the effects of norms treatment versus control on petition signing. A sample of about 364 students would achieve a power of 0.8 with an alpha of 0.05. With our final sample size of 560 students (370 between norms treatment and control), the study is sufficiently powered at 0.8, exceeding the typical threshold for adequate hypothesis testing.[16]

Fig. 5 shows that there are no differences in signing the petition across treatment groups in the online experiment. This finding underscores the plausibility of our interpretation of the main findings; namely, that the treatment effect passes through actual compliance.

6. Mechanisms

What explains the positive effect of ticket payment — induced by priming cooperative norms — on the likelihood of signing a petition for quality public transportation? Based on our characterization of individuals influenced by this intervention (homo reciprocans), the key mechanism is the intervention's solution to the collective action dilemma. By signaling that other passengers contribute to the public good and are thus likely to sign the petition, the message fosters a sense of collective efficacy, reassuring individuals that their participation is part of a broader cooperative effort.

However, this assurance applies only to horizontal reciprocity — trust among passengers. For the petition to be successful, homo reciprocans must also believe in vertical reciprocity — trust in institutions to respond to citizen demands. As demonstrated in Section 5.4, merely priming collective efficacy without actual compliance is insufficient to drive accountability-seeking behavior.

As proposed in Section 3, actual compliance is crucial because it fosters a sense of entitlement—paying for a ticket reinforces individuals' legitimacy in demanding quality service. When compliance is driven

---

[15] We acknowledge that some subjects may decide to not sign the animal rights' petition because they do not agree with its content; not necessarily because they are not prone to participate in a relatively costly collective action (remember that petition signatories must provide their name and surname). However, the proportion of subjects who disagrees with the petition's content should be the same across all experimental groups (due to randomization), still allowing us to parse out a potential shift in the willingness to sign the falsification petition due to priming of collective efficacy in the norms treatment.

[16] Figure A8 in the Appendix visualizes the power across different sample sizes, reaffirming the robustness of our sample size in detecting meaningful effects.

Fig. 4. Treatment messages conveyed as a preview of lottery prizes.
Notes: The figure shows the T-shirts that participants of the follow-up survey could win in a lottery. They served as our informational treatments in the online experiment. The upper-left T-shirt shows the control message; the upper-right T-shirt shows the T1 (sanctions) message; the bottom T-shirt shows the T2 (norms) message.
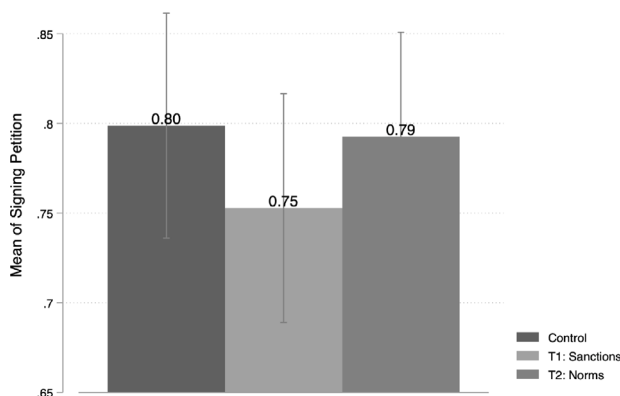

Fig. 5. Falsification test in the online experiment.
Notes: The figure shows the comparison of means of signing the petition across treatment conditions in the follow-up online experiment. Respondents were recruited among students of the University of Buenos Aires.

by an appeal to broad cooperative norms, this entitlement is likely to align with expectations of collective commitment to accountability-seeking—namely, the belief that others will also sign the petition for the same reason. This perceived collective effort, in turn, serves as a credibility signal to policymakers, increasing the likelihood of their response and thus reinforcing the subjects' expectations of institutional responsiveness.

Below, we present descriptive evidence suggesting that these mechanisms play a role. Specifically, in the final three weeks of our experiment, we measured perceptions of the right to demand accountability, while expectations of state reciprocity — proxied by trust in government — were assessed during the initial eight weeks.[17] Table 6 shows that the norms-based intervention increased perceptions of one's right to demand accountability (columns 1 and 3) and enhanced trust in

government among compliant citizens (columns 4 and 6). This pattern is consistent across both the full IV regressions and reduced-form models. In contrast, the sanctions-based intervention did not produce similar effects (columns 2 and 5).

Additional implications. Last, we explore two additional implications of the aforementioned mechanism. However, these additional analyses are constrained by the limited data we were able to gather during our experiment. It is important to note that we faced significant time constraints, as we had only a brief window to conduct interviews with passengers on the platform.

First, if our mechanism is correct, we might expect that the normative channel should produce larger effects among women, compared to men. Women have been consistently shown to exhibit more egalitarian preferences and be more worried about fairness than men (Croson and Gneezy, 2009; Mutz and Lee, 2020). The related evidence spans many fields of social sciences, from psychology (Gilligan, 1982), through sociology (Eckel and Grossman, 2001), to decision/management science (Babcock and Laschever, 2003). Based on these documented gender differences with respect to the cooperative "type", we thus expect that women should respond to the "compliance-accountability" logic more promptly (Bates and Lien, 1985). In line with this prediction, in Fig. 6, we find that compliance induced by the priming of cooperative norms indeed increases accountability-seeking among women, but not among men. This pattern is therefore consistent with the proposed mechanism.

Second, if our mechanism is correct, we should expect that the reported effect of compliance on individual contribution to accountability-seeking is weaker in poorer neighborhoods. The relative poverty of these areas may serve as a tangible sign of the government's failure to enhance citizens' well-being, undermining the expectation of vertical reciprocity from the state. We measure neighborhood's wealth with indicators of night-time luminosity and population density in the area surrounding a given metro-train station. The results are presented in Figures A6 and A7, confirming that the reported effects of normative appeals are concentrated in wealthier neighborhoods.

How much of this "type of neighborhood" effect is driven by the compositional differences between neighborhood populations and how much is explained by the signals of government's responsiveness to

---

[17] We could not measure all outcomes simultaneously due to the limited time enumerators had to interview passengers on the platform.
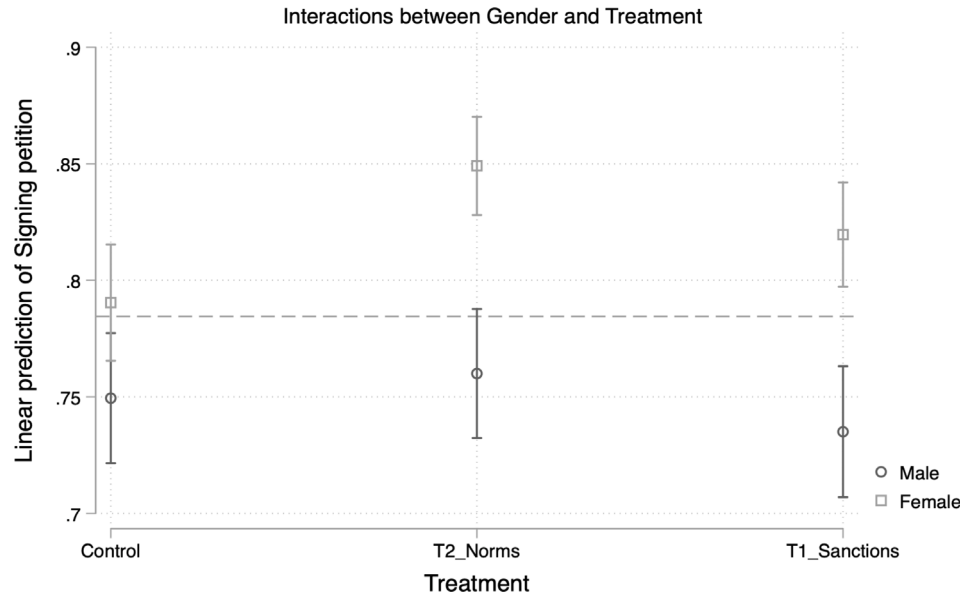
Fig. 6. Treatment, compliance, and signing the petition (heterogeneity along gender).
Notes: The figure shows the point-estimates and the accompanying 90/95 confidence intervals (thick and thin lines, respectively) of the regression of signing the petition on treatment conditions conditional on respondents' gender. The dashed line indicates the mean level of petition signing for the whole sample across all treatment conditions.

Table 6
Treatment, compliance, and the perceived right to the quality public service.

	(1)	(2)	(3)	(4)	(5)	(6)
	Right to quality public service			Trust in the government		
	Reduced form	IV = T1 Sanctions	IV = T2 Norms	Reduced form	IV = T1 Sanctions	IV = T2 Norms
T1 (Sanctions)	0.122** (0.050)			0.066 (0.101)		
T2 (Norms)	0.160*** (0.050)			0.231** (0.100)		
Paid ticket		3.090 (3.095)	1.899* (1.109)		0.400 (0.828)	4.077* (1.951)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	673	435	418	4268	2831	2822

Notes: Robust standard errors in parenthesis. Significance level shown below *p<0.10, ** p<0.05, ***p<0.01.

citizens' needs? A useful approach to address this question is proposed by Angrist et al. (2013), who employ Blinder-Oaxaca decomposition of the Local Average Treatment Effect (LATE). To formalize this notion in our case study, consider the Blinder-Oaxaca decomposition of the LATE difference between wealthier and poorer neighborhoods. Let τ_w represent the LATE for norms (vis-á-vis sanctions) interventions in wealthier neighborhoods and τ_p represent the LATE in poorer neighborhoods. The decomposition can be expressed as:

$$\tau_w - \tau_p = (\mu_w - \mu_p) + \overline{X}(\rho_w - \rho_p) + \rho(\overline{X}_w - \overline{X}_p) \tag{5}$$

Whereby μ_w and μ_p are the mean outcomes for compliers under norms (vis-á-vis sanctions) interventions in wealthier and poorer neighborhoods, respectively. $\overline{X}_w$ and $\overline{X}_p$ are the average values of selected covariates for the compliers in each group. ρ_w and ρ_p are the coefficients associated with these covariates for the respective groups, while ρ represents the overall effect of these covariates.

The term $\rho(\overline{X}_w - \overline{X}_p)$ captures the part of the accountability-seeking gap explained by demographic differences between compliers in the two groups. The terms $(\mu_w - \mu_p)$ and $\overline{X}(\rho_w - \rho_p)$ capture the components of the gap attributable to effect heterogeneity within demographic groups. By incorporating the Blinder-Oaxaca decomposition, we can thus better dissect the underlying demographic and contextual factors influencing compliance behavior. To implement the Blinder-Oaxaca decomposition to analyze our data, we use -oaxaca- command in Stata

developed by Jann (2008). We divide neighborhoods into wealthier and poorer ones around the median of night-time luminosity. Following Lin (2013), we estimate a fully interacted model, i.e., interacting treatment with each covariate.

Table 7 shows the result of the decomposition analyses. The mean of the signing the petition is 0.77 for people in poorer neighborhoods and 0.81 for people in wealthier neighborhoods, yielding an accountability seeking gap of 0.17. In the second panel of the decomposition output, the accountability-seeking gap is divided into two parts. The first part reflects the mean increase in accountability-seeking if poorer neighborhoods had the same characteristics as the wealthier ones. The increase of 0.012 indicates that differences in endowments do not account for the observed gap; to the contrary, they attenuate the gap (note the opposite signs of the "Explained" and "Difference" coefficients).

The second term in the decomposition panel of Table 7 quantifies the change in the neighborhood's population signing the petition when applying the wealthier neighborhoods's coefficients to the poorer neighborhood's effect of our intervention. This is the "unexplained" part of the LATE difference. The large negative coefficient indicates that the compositional differences and their potential moderation of our treatment effect (captured by interactions) do not explain the gap between the types of neighborhoods, lending further support to our interpretation of the heterogeneous effects.

Table 7

Blinder-Oaxaca Decomposition of LATE (norms vs sanctions interventions)

	Coef.	Std. Err.
Differential		
Prediction (poorer areas)	0.7675	(0.0126)
Prediction (wealthier areas)	0.8120	(0.0092)
Difference	−0.0445	(0.0156)
Decomposition		
Explained	0.0122	(0.0039)
Unexplained	−0.0567	(0.0155)

7. Conclusion

In this study, we conducted a field experiment at train stations in Buenos Aires to estimate the effect of civic compliance on accountability-seeking behaviors. We found that appealing to the fear of sanctions and priming cooperative norms increase compliance with paying the public transportation ticket. However, only appeals to norms also make passengers more willing to sign the petition demanding quality public services—an important indication of their propensity to contribute to the collective action of holding their governors accountable. Our findings suggest that raising compliance through appeals to norms may have wider societal benefits, compared to interventions inducing compliance through the threat of sanctions. These results are broadly consistent with the proposition that compliance-inducing interventions may have spillover effects to other domains of civic life (Altmann et al., 2021), while offering important qualifications to this established pattern.

Our results point to heterogeneities in the effects of compliance-inducing interventions. We argue that different types of individuals might respond to each of our treatments. Namely, individuals who fit the *homo-economicus* type are likely to change their propensity to pay the ticket when exposed to the sanctions treatment because the expected costs of compliance outweigh the expected costs of penalties for non-compliance. Yet, homo economicus are free-riders and thus their choice to pay the ticket does not affect their likelihood to engage in other forms of collective actions, such as contributions to the provision of the public good of accountability. By contrast, individuals who react to the invocation of norms are likely to be the *homo-reciprocans* type (Fehr and Gächter, 1998). They demand quality public services after paying the transportation fare because they follow the logic of vertical (citizen-state) reciprocity of this exchange—rights in return for duties (and vice versa). They may also recognize collective responsibilities to contribute to public goods, as captured by the notions of horizontal reciprocity (citizen-citizen), conceptualizing accountability-seeking as an extension of their normative obligation to comply with civic duties if others do the same.

Our key contribution is micro evidence on the "taxation produces representation" hypothesis and its underlying mechanism (Peruzzotti and Smulovitz, 2006). Our study provides rare causal evidence supporting this proposition with reference to the domain of everyday compliance. A variant of this hypothesis that focuses on voting behavior have been studied with the use of quasi-experimental methods (see, e.g., Paler, 2013; Weigel, 2020). However, everyday compliance differs from the election setting insofar as it involves *direct* exchange of payment for services, which plausibly lowers citizens' urge to monitor what the government does *after* it had already delivered the services they paid for. Our study thus offers a 'hard' test on the "compliance breeds accountability" hypothesis.

We are aware that compliance with civic duties and engaging in accountability-seeking behaviors include many actions that are certainly costlier than paying a train ticket or signing a petition. However, we focused on this outcome because it is malleable to weak informational treatments. Our study thus provides evidence-based policy recommendation on cheap and potentially scalable ways of improving everyday forms of compliance and their effect on civicness at large.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2025.103492.

Data availability

Data will be made available on request.

References

Adida, Claire, Gottlieb, Jessica, Kramon, Eric, McClendon, Gwyneth, 2020. When does information influence voters? The joint importance of salience and coordination. Comp. Political Stud. 53 (6), 851–891.

Alm, James, 2012. Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. Int. Tax Public Financ. 19 (1), 54–77.

Alm, James, Bloomquist, Kim M., McKee, Michael, 2017. When you know your neighbour pays taxes: Information, peer effects and tax compliance. Fisc. Stud. 38 (4), 587–613.

Alm, James, Kirchler, Erich, Muehlbacher, Stephan, Gangl, Katharina, Hofmann, Eva, Kogler, Christoph, Pollai, Maria, 2012. Rethinking the research paradigms for analysing tax compliance behaviour. In: CESifo Forum. ifo Institut-Leibniz-Institut, 13 München, pp. 33–40.

Altmann, Steffen, Grunewald, Andreas, Radbruch, Jonas, 2021. Altmann steffen andreas grunewald jonas radbruch interventions and cognitive spillovers. Review of Economic Studies.

Andreoni, James, Erard, Brian, Feinstein, Jonathan, 1998. Tax compliance. J. Econ. Lit. 36 (2), 818–860.

Andriani, Luca, 2016. Tax morale and prosocial behaviour: Evidence from a Palestinian survey. Camb. J. Econ. 40 (3), 821–841.

Andrighetto, Giulia, Zhang, Nan, Ottone, Stefania, Ponzano, Ferruccio, D'Attoma, John, Steinmo, Sven, 2016. Are some countries more honest than others? Evidence from a tax compliance experiment in Sweden and Italy. Front. Psychol. 7, 472.

Angrist, Joshua D., Pathak, Parag A., Walters, Christopher R., 2013. Explaining charter school effectiveness. Am. Econ. Journal: Appl. Econ. 5 (4), 1–27.

Armand, Alex, Augsburg, Britta, Bancalari, Antonella, 2021. Public service delivery and free riding: Experimental evidence from India. CEPR Discussion Paper No. DP16284.

Babcock, Linda, Laschever, Sara, 2003. Women Don'T Ask: Negotiation and the Gender Divide. Princeton University Press.

Bablawi, Hazem, Luciani, Giacomo, 1987. The Rentier State (Nation, State and Integration in the Arab World). Routledge Kegan & Paul, London.

Bates, Robert H., Lien, Da-Hsiang Donald, 1985. A note on taxation, development, and representative government. Politics & Soc. 14 (1), 53–70.

Bicchieri, Cristina, 2005. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, Cambridge.

Bursztyn, Leonardo, Fiorin, Stefano, Gottlieb, Daniel, Kanz, Martin, 2019. Moral incentives in credit card debt repayment: Evidence from a field experiment. J. Political Econ. 127 (4), 1641–1683.

Castro, Lucio, Scartascini, Carlos, 2015. Tax compliance and enforcement in the pampas: Evidence from a field experiment. J. Econ. Behav. Organ. 116, 65–82.

Chan, Ho Fai, Supriyadi, Mohammad Wangsit, Torgler, Benno, 2018. Trust and tax morale. In: The Oxford Handbook of Social and Political Trust. Oxford University Press, New York, pp. 1–71.

Chaudhry, Kiren Aziz, 1997. The Price of Wealth: Economies and Institutions in the Middle East. Cornell University Press, Ithaca.

Coleman, Stephen, 1996. The Minnesota income tax compliance experiment: State tax results. Minnesota Department of Revenue.

Croson, Rachel, Gneezy, Uri, 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474.

Dai, Zhixin, Galeotti, Fabio, Villeval, Marie Claire, 2018. Cheating in the lab predicts fraud in the field: An experiment in public transportation. Manag. Sci. 64 (3), 1081–1100.

Daude, Christian, Gutiérrez, Hamlet, Melguizo, Ángel, 2012. What Drives Tax Morale? Organisation for Economic Cooperation and Development.

Dickson, Eric S., Gordon, Sanford C., Huber, Gregory A., 2022. Identifying legitimacy: experimental evidence on compliance with authority. Sci. Adv. 8 (7).

Dohmen, Thomas, Falk, Armin, Huffman, David, Sunde, Uwe, 2009. Homo reciprocans: Survey evidence on behavioural outcomes. Econ. J. 119 (536), 592–612, eabj7377.

Dularif, Muh, Rustiarini, Ni Wayan, 2021. Tax compliance and non-deterrence approach: a systematic review. Int. J. Sociol. Soc. Policy.

Eckel, Catherine C., Grossman, Philip J., 2001. Chivalry and solidarity in ultimatum games. Econ. Inq. 39 (2), 171–188.

Elster, Jon, 1999. Alchemies of the Mind: Rationality and the Emotions. Cambridge University Press, Cambridge.

Falk, Armin, Fischbacher, Urs, 2006. A theory of reciprocity. Games Econom. Behav. 54 (2), 293–315.

Fehr, Ernst, Gächter, Simon, 1998. Reciprocity and economics: The economic implications of homo reciprocans. Eur. Econ. Rev. 42 (3–5), 845–859.

Fehr, Ernst, Gächter, Simon, 2000a. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.

Fehr, Ernst, Gächter, Simon, 2000b. Fairness and retaliation: The economics of reciprocity. J. Econ. Perspect. 14 (3), 159–182.

Fehr, Ernst, Gächter, Simon, 2002. Altruistic punishment in humans. Nature 415 (6868), 137–140.

Fergusson, Leopoldo, Molina, Carlos, Riaño, Juan.Felipe, 2019. Consumers as VAT evaders. Economía 19 (2), 21–68.

Frey, Bruno S., Torgler, Benno, 2007. Tax morale and conditional cooperation. J. Comp. Econ. 35 (1), 136–159.

Gadenne, Lucie, 2017. Tax me, but spend wisely? Sources of public finance and government accountability, American economic. J.: Appl. Econ. 274–314.

Gelman, Andrew, Hill, Jennifer, Vehtari, Aki, 2021. Regression and Other Stories. Cambridge University Press, Cambridge.

Gilligan, Carol, 1982. In a Different Voice: Psychological Theory and Women's Development. Harvard University Press.

Gonzalez-Navarro, Marco, Quintana-Domeque, Climent, 2013. Local Public Goods and Property Tax Compliance: Evidence from Residential Street Pavement. Lincoln Institute of Land Policy.

Hallsworth, Michael, 2014. The use of field experiments to increase tax compliance. Oxf. Rev. Econ. Policy 30 (4), 658–679.

Hallsworth, Michael, List, John A., Metcalfe, Robert D., Vlaev, Ivo, 2017. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. J. Public Econ. 148, 14–31.

Heckman, James J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In: Annals of Economic and Social Measurement, vol. 5, NBER, pp. 475–492.

Hodgson, Geoffrey M., 2012. From Pleasure Machines To Moral Communities: An Evolutionary Economics Without Homo Economicus. University of Chicago Press.

Holland, Alisha C., 2015. The distributive politics of enforcement. Am. J. Political Sci. 59 (2), 357–371.

Holland, Alisha C., 2016. Forbearance. Am. Political Sci. Rev. 110 (2), 232–246.

Horodnic, Ioana Alexandra, 2018. Tax morale and institutional theory: A systematic review. Int. J. Sociol. Soc. Policy.

Ibrahim, Muazu, Musah, Alhassan, Abdul-Hanan, Abdallah, 2015. Beyond enforcement: What drives tax morale in Ghana? Humanomics.

Jann, Ben, 2008. A stata implementation of the blinder-oaxaca decomposition. Stata J. 8 (4), 453–479.

Katz, Barbara G., Owen, Joel, 2013. Exploring tax evasion in the context of political uncertainty. Econ. Syst. 37 (2), 141–154.

Kirchler, Erich, Hoelzl, Erik, Wahl, Ingrid, 2008. Enforced versus voluntary tax compliance: The slippery slope framework. J. Econ. Psychol. 29 (2), 210–225.

Kirchler, Erich, Muehlbacher, Stephan, Kastlunger, Barbara, Wahl, Ingrid, et al., 2010. Why pay taxes? A review of tax compliance decisions. In: Developing Alternative Frameworks for Explaining Tax Compliance, vol. 59.

Kondelaji, MirHadi Hosseini, Sameti, Majid, Amiri, Hadi, Moayedfar, Rozita, 2016. Analyzing determinants of tax morale based on social psychology theory: Case study of Iran. Iran. Econ. Rev. 20 (4), 581–598.

Leonardo, Gabriel Marcelo, 2011. Politics and Tax Morale. the Role of Trust, Values, and Beliefs in Shaping Individual Attitudes Towards Tax Compliance. Georgia Institute of Technology, PhD Dissertation.

Levi, Margaret, Stoker, Laura, 2000. Political trust and trustworthiness. Annu. Rev. Political Sci. 3 (1), 475–507.

Lin, Winston, 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. Ann. Appl. Stat. 7 (1), 295–318.

Lindenberg, Siegwart, Six, Frederique, Keizer, Kees, 2021. Social contagion and goal framing: The sustainability of rule compliance. In: The Cambridge Handbook of Compliance. Cambridge University Press, Cambridge, pp. 422–437.

Mascagni, Giulia, 2018. From the lab to the field: A review of tax experiments. J. Econ. Surv. 32 (2), 273–301.

Matikinca, Phikolomzi, Ziervogel, Gina, Enqvist, Johan.P, 2020. Drought response impacts on household water use practices in Cape Town. South Afr. Water Policy 22 (3), 483–500.

Mogstad, Magne, Torgovitsky, Alexander, Walters, Christopher.R, 2021. The causal interpretation of two-stage least squares with multiple instrumental variables. Am. Econ. Rev. 111 (11), 3663–3698.

Mogstad, Magne, Torgovitsky, Alexander, Walters, Christopher.R, 2024. Policy evaluation with multiple instrumental variables. J. Econometrics 105718.

Murphy, Ryan O., Ackermann, Kurt A., Handgraaf, Michel J.J., 2011. Measuring social value orientation. Judgm. Decis. Mak. 6 (8), 771–781.

Mutz, Diana C., Lee, Amber Hye-Yon, 2020. How much is one American worth? How competition affects trade preferences. Am. Political Sci. Rev. 114 (4), 1179–1194.

Nagel, Rosemarie, 1995. Unraveling in guessing games: An experimental study. Am. Econ. Rev. 85 (5), 1313–1326.

North, Douglass C., Weingast, Barry R., 1989. Constitutions and commitment: The evolution of institutions governing public choice in seventeenth-century England. J. Econ. Hist. 49 (4), 803–832.

Nurkholis, Nurkholis, Dularif, Muh, Rustiarini, N .Wayan, 2020. Tax evasion and service-trust paradigm: A meta-analysis. Cogent Bus. Manag. 7 (1), 1827699.

Ortega, Daniel, Ronconi, Lucas, Sanguinetti, Pablo, 2016. Reciprocity and willingness to pay taxes: Evidence from a survey experiment in latin america. Economía 16 (2), 55–87.

Paler, Laura, 2013. Keeping the public purse: An experiment in windfalls, taxes, and the incentives to restrain government. Am. Political Sci. Rev. 107 (4), 706–725.

Peruzzotti, Enrique, Smulovitz, Catalina, 2006. Enforcing the Rule of Law: Social Accountability in the New Latin American Democracies. University of Pittsburgh Press, Pittsburgh.

Porumbescu, Gregory A., Lindeman, Meghan I.H., Ceka, Erica, Cucciniello, Maria, 2017. Can transparency foster more understanding and compliant citizens? Public Adm. Rev. 77 (6), 840–850.

Pruckner, Gerald J., Sausgruber, Rupert, 2013. Honesty on the streets: A field study on newspaper purchasing. J. Eur. Econ. Assoc. 11 (3), 661–679.

Riahi-Belkaoui, Ahmed, 2004. Relationship between tax compliance internationally and selected determinants of tax morale. J. Int. Account. Audit. Tax. 13 (2), 135–143.

Ritsatos, Titos, Tax evasion and compliance from the neo classical paradigm to behavioural economics: A review. J. Account. Organ. Chang. 2014.

Ronconi, Lucas, 2019. From citizen's rights to civic responsibilities. IZA Discussion Paper.

Ross, Michael L., 2004. Does taxation lead to representation? Br. J. Political Sci. 34 (2), 229–249.

Saulitis, Andris, 2023. Nudging debtors with non-performing loans: Evidence from three field experiments. J. Behav. Exp. Financ. 37, 100776.

Slemrod, Joel, Blumenthal, Marsha, Christian, Charles, 2001. Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in minnesota. J. Public Econ. 79 (3), 455–483.

Sliwka, Dirk, 2007. Trust as a signal of a social norm and the hidden costs of incentive schemes. Am. Econ. Rev. 97 (3), 999–1012.

Steinmo, Sven, D'Attoma, John, 2021. Steinmo Sven John D'Attoma Willing To Pay?: A Reasonable Choice Approach. Oxford University Press, Oxford.

Szekely, Aron, Lipari, Francesca, Antonioni, Alberto, Paolucci, Mario, Sánchez, Angel, Tummolini, Luca, Andrighetto, Giulia, 2021. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. Nat. Commun. 12 (1), 5452.

Tilly, Charles, 1992. Coercion, Capital, and European States, AD 990-1992. Blackwell, Oxford.

Torgler, Benno, 2003. Tax morale, rule-governed behaviour and trust. Const. Political Econ. 14 (2), 119–140.

Torgler, Benno, Werner, Jan, 2005. Fiscal autonomy and tax morale: Evidence from Germany. CREMA Working Paper.

Waterbury, John, 1997. Fortuitous by-products. Comp. Politics 383–402.

Weigel, Jonathan L., 2020. The participation dividend of taxation: How citizens in congo engage more with the state when it tries to tax them. Q. J. Econ. 135 (4), 1849–1903.