WORKING PAPER Nº IDB-WP-1546

Transparency and Government Reputation:

An Experiment on Signaling

Susana Otálvaro-Ramírez
Carlos Scartascini
Jorge M. Streb

Inter-American Development Bank
Department of Research and Chief Economist

February 2025

Transparency and Government Reputation:

An Experiment on Signaling

Susana Otálvaro-Ramírez*
Carlos Scartascini**
Jorge M. Streb***

* University of California, Santa Barbara
** Inter-American Development Bank
*** Universidad del CEMA

Inter-American Development Bank
Department of Research and Chief Economist

February 2025

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Otálvaro-Ramírez, Susana.
Transparency and government reputation: an experiment on signaling / Susana Otálvaro-Ramírez, Carlos Scartascini, Jorge M. Streb.
p. cm. — (IDB Working Paper Series ; 1546)
Includes bibliographical references.
1. Transparency in government-Argentina. 2. Trust-Political aspects-Argentina. 3. Public administration-Argentina I. Streb, Jorge Miguel. II. Scartascini, Carlos G., 1971- III. Inter-America Development Bank. Department of Research and Chief Economist. IV. Title. V. Series.
IDB-WP-1546

http://www.iadb.org

Abstract

Transparency initiatives are well-known tools to foster trust and empower citizens. To explain why some governments introduce them but others do not, we model these initiatives as a signal that complements the information provided by visible government performance and conduct a randomized survey experiment in the City of Buenos Aires, Argentina, where the incumbent mayor made a set of post-electoral promises. In a setting with relatively high trust priors, our results show that these initiatives matter in shaping citizens' perceptions of the reputation of the government. We find, however, strong heterogeneity among three groups of citizens. A group unfamiliar with the policy was impervious to treatment: they seem to react to deeds, not words, and have, on average, lower initial trust. The treatment effects are entirely through those vaguely familiar with the promises, closing the average gap in trust with those familiar with the promises. More generally, our study suggests that transparency initiatives may be an effective signal, though their informational value may be more limited than visible public performance.

JEL classifications: D72, D78, D82, D83, H41

Keywords: Rent-seeking, Promises, Signaling, Social learning, Trust, Reputation, Political economy, Development, Survey experiment

We are grateful to Martin Alessandro and Jerónimo Torrealday for inviting us to evaluate this policy and leading the work while at the city government, and to the many people who commented on earlier versions of this paper during presentations at UPenn, CIDE, Hertie School, ITAM, MPSA, EJPE at GMU, LACEA, and the Public Economics RIDGE Forum. This research was financed by the Inter-American Development Bank (IDB). IDB management had no involvement in the study design, analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the article for publication. The opinions expressed here are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent. Data collection was approved by the Government of the City of Buenos Aires, and complies with all the ethics regulations of the City of Buenos Aires Government and the Inter-American Development Bank. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

1 Introduction

Access to information allows citizens to hold governments accountable (Khemani et al., 2016) and bolsters their power to reward or censure elected officials for their performance in office (Ferraz and Finan, 2008; Kendall et al., 2015). In most cases, transparency is exogenously introduced. For example, the IMF or the OECD push reforms to their members. In other cases, higher levels of government introduce monitoring mechanisms for lower ones. An important question is under what conditions policymakers have the incentive to introduce transparency initiatives that constrain them and what effects those initiatives may have on the political equilibrium.

Theoretically, we conceptualize transparency initiatives as a signal of the trustworthiness of incumbent governments that can improve their reputation. Transparency initiatives allow honesty to have a bite insofar as this signal is costlier if a government must cover up what it is actually doing. Moreover, these initiatives can make it less costly for benevolent governments to signal their type than through the misallocation of resources from less visible to more visible public goods. We develop an analytical framework where a benevolent government will always be willing to launch a reform that increases transparency, while the choice of non-benevolent ones will depend on the cover-up costs of acting dishonestly. Since trustworthiness is a valence issue, incumbents who are seen as more trustworthy (have a better reputation) will have an advantage over challengers (see Ansolabehere and Snyder, 2000).

The empirical analysis is based on a survey experiment carried out in the City of Buenos Aires, Argentina, a developing country, within an actual policy setting. In this survey experiment, we provided information about the government's post-electoral promises and performance. The treated and control individuals were asked about their perceptions regarding the city government's degree of trustworthiness using a multidimensional approach that includes the components of trust listed by Grimmelikhuijsen (2012): competence, benevolence, and honesty.[1] This approach helps us shed light on how transparency initiatives can influence the government's reputation among citizens.

Does, in fact, more information improve an incumbent's reputation? While there are examples of successful transparency initiatives (Alessandro et al., 2021; Ardanaz et al., 2023), evidence of

[1] See also Hamm et al. (2019), who follow the organizational model of trust developed by Schoorman et al. (2007) to understand the psychological nature and mechanisms of trust, and identify the integrity, ability, and benevolence of governments as mediators of information on final trust perceptions update. Trust in government differs from trust in government members (Keefer et al., 2018, 2022).

the particular conditions under which transparency fosters trust is scant, especially in the context of developing countries and young democracies (Blanco and Ruiz, 2013; Evans et al., 2019). Our framework leads us to explore three factors that affect the impact of transparency initiatives and, hence, whether incumbents may be willing to introduce them: the content of the information provided, the priors citizens have about the government, and citizens' reaction to socially transmitted information.

First, transparency initiatives' effect can differ because of their informational content.[2] Participants were randomly assigned to either a control group, which did not receive any information, or one of three informational treatments. Treatment 1 (T^1) provides information about a series of pledges made by the Mayor of Buenos Aires at the beginning of the government period. Treatments 2 (T^2) and 3 (T^3) supplement the general information provided in Treatment 1 by providing information about compliance with the pledges at the aggregate (city), or local level (commune - *comuna*).[3] This information is not provided in a vacuum; citizens can compare it to their personal experiences and make assessments of the government based on that comparison.

A second reason for differential impacts is the two-way relationship between trust and transparency: with Bayesian updating, the treatment's impact on posteriors, i.e., final trust, will depend on priors, i.e., initial trust. Since most people from the control group consider the government quite trustworthy, we expect that treatments will have positive and significant effects on reputation ($T^1 > 0$, $T^2 > 0$, $T^3 > 0$). Disclosing information indeed increases trust in government by about 0.1 standard deviations (SD). Though informing about the compliance with the commitments should have a more substantial effect than informing only about the existence of the commitments ($T^2 \geq T^1$, $T^3 \geq T^1$), effects tend to be relatively uniform across the different components of our main index. The coefficients for T^2 and T^3 are positive and statistically significant, and they tend to be larger than those of T^1 but are not statistically different from it.

Average effects, however, present an incomplete account of how treatments interact with be-

[2] Alessandro et al. (2021), for instance, show that people who received a treatment indicating that the government was over-performing on its promises trusted the government more than those who received a treatment showing that the government was under-performing. If there is nothing good to show, more transparency might even hurt the incumbent (Piotrowski et al., 2019).

[3] Since dependent variables are measured shortly after the treatment, there is potential for interviewer demand effect. However, results are not uniform across the three treatments and dependent variables, which provides reassurance that respondents did not simply rate the government more favorably after getting the treatment due to a bias, but in response to the information provided in the vignettes.

liefs because respondents have heterogeneous priors. We expect the most substantial effect of the transparency initiative to be on those with intermediate trust priors, because those who do not trust the government at all should be unaffected by the treatments, and those with the highest trust face a ceiling effect. We examine this prediction by looking at how trust changes along the full spectrum of priors, as well as the effect of treatments conditional on the perceived quality of government, a variable closely related to initial trust. People with intermediate priors are indeed the most responsive to information.

A third reason for differential impacts is that citizens have heterogeneous learning behaviors, as revealed by differences in their starting knowledge of city government plans and actions. Since the information set of the group already aware of the government's pledges does not change with the treatment, information should only affect individuals not familiar with them. While the data confirm this prediction, we did not foresee the existence of two very distinct groups among people not wholly familiar with the government's plans. A group vaguely familiar with the post-electoral promises, i.e., those who had heard about it, reacted strongly to the treatment assignment. A second group, completely unaware of the promises, showed no response at all to the informational treatment. While we cannot prove causality in our setting, there is suggestive evidence that this group of people does not rely on others (family, social, or traditional media) nor on the government to acquire information about the government's performance.[4]

2 Analytical Framework

The transparency initiative we study provides information on plans as well as on performance. The Mayor of Buenos Aires referred to these post-electoral promises as "*compromisos,*" i.e., commitments.[5] Why may transparency initiatives like this matter? Our explanation is that they may act as a signal about government type that complements other information voters already have on government performance, like, for example, the state of the economy (Sances, 2021). In our experiment, respondents live in the city, so they can rely on their own experience to see whether

[4]The absence of an impact on this group can also be attributed to the smaller sample size. We will later go into greater detail about the absence of effect.

[5]These plans are available at http://www.buenosaires.gob.ar/compromisos. For example, the Mayor of CABA said in a public statement after he was elected: "We are committed to ensuring that, during this term of office, 20,000 families will be able to fulfill their dream of owning their home." Another example was the building of eight educational centers, mainly located in vulnerable neighborhoods of the city.

the government actually keeps its promises or not.

We incorporate transparency initiatives into a model where the incumbent has two signals: visible public goods and a transparency initiative that can increase information on the provision of less visible public goods. Both signals are analyzed jointly. We then discuss how trust priors affect the way in which people process this verbal information, in a setup where there is significant heterogeneity among respondents, rendering the common prior assumption invalid. Finally, the most noteworthy feature in our experiment is that respondents vary greatly in their familiarity with the transparency initiative. We interpret this as the result of heterogeneity in how much individuals trust social information.

2.1 Government Signals

We set up a workhorse model to analyze how the information provided by transparency initiatives can affect the reputation of the government. We integrate transparency initiatives into the framework provided by the literature on rational retrospective voting that starts with Rogoff and Sibert (1988), which shows that information on government performance matters.[6]

We develop a setting where benevolence, competence, and honesty go together naturally: a benevolent government does not divert resources to its own pockets, so it can provide more public goods with a given budget, being perceived as more competent; since it has nothing to hide, it can also be open and honest about what it does.[7] In our two-period model, incumbents that gain the reputation that they are trustworthy have higher chances of reelection, so non-benevolent incumbents are tempted to send the same signal as benevolent governments in the first period.

We model the transparency initiative as a costly signal that can potentially reveal information about government goals and performance, because some dishonest government types are deterred from sending it due to cover-up costs.[8] As we will discuss in the empirical section, the transparency reform positively impacts respondents in the treatment group, who trust the government more and see it as more benevolent, honest, and competent.

[6] Ashworth (2012) summarizes theoretical and empirical literature to support this fact.

[7] This reflects the feature that the responses about trustworthiness and its three dimensions (benevolence, honesty, and competence) are driven in the sample by a single factor according to parallel and factor analyses.

[8] Since post-electoral promises involve an administrative reform that can improve the monitoring of the public administration, this might enhance the provision of public goods. We abstract from this issue in what follows.

Visible Public Goods

The basic signal to voters is the provision of visible public goods. This setup characterizes the control group, those not exposed to the post-electoral promises. Visible public goods g_v, which go from varieties 0 to v, are immediately observed by the voter. Non-visible public goods g_{nv}, which go from v to 1, are observed later. The per-period utility u is concave in the consumption of public goods. We further assume that utility is logarithmic in consumption, $u(.) = \ln(.)$, leading to an explicit analytical solution. Since utility is the same for each variety, incumbents will want to provide the same quantity within each group:

$$u(g_{vt}, g_{nvt}) = v \ln(g_{vt}) + (1 - v) \ln(g_{nvt}).$$

Each individual citizen i is subject to a shock σ_i that has an idiosyncratic component identically distributed over time. Instead of a shock to preferences, as in the random utility model (McFadden, 1981), we treat σ_i as imprecise perceptions of public good provision (Woodford, 2020). We assume the shock is uniformly distributed around zero, so the median voter $i = m$ is not affected by the idiosyncratic shock.[9]

$$u_{it} = u(g_{vt}, g_{nvt}) + \ln(1 + \sigma_{it}).$$

In the model, the perception of the provision of public goods determines the incumbent's reputation, so this shock can capture the heterogeneous perceptions of trustworthiness in the data.[10] A voter's expected utility is given by the discounted sum $\mathbb{E}_t[U_{it}] = \mathbb{E}_t[\sum_{t=1}^{2} \delta^{t-1} u_{it}]$.

Incumbents have the same preferences as citizens, as in the citizen-candidate models (Besley and Coate, 1997). They differ in benevolence, which is captured by whether the per-period utility u of the incumbent depends or not on an additional term r of personal rents:

$$u_{jt} = u(g_{vt}, g_{nvt}) + \alpha_j r_t,$$

[9] This makes voting deterministic, because candidates have complete information about the median voter that is decisive in choosing between the incumbent and the challenger. If there were also a common component in the political shock, it would make voting probabilistic.

[10] Figures A3 to A5 and Figure A8 show heterogeneity of priors for the control group.

where a benevolent government, $j = b$, has $\alpha_b = 0$, while a non-benevolent government, $j = nb$, has $\alpha_{nb} = \alpha > \underline{\alpha} > 0$, where $\underline{\alpha}$ is the threshold level beyond which personal rents are positive, as shown below. The expected utility of the incumbent is given by $\mathbb{E}_t[U_{jt}] = \mathbb{E}_t[\sum_{t=1}^{2} \delta^{t-1} u_{jt}]$.

Public expenditure determines the provision of public goods, i.e.,

$$g_{st} = \gamma_{st} \quad \text{for} \quad s = v, nv.$$

The per-period budget constraint is that government expenditures γ_s, for $s = v, nv$, plus rents r appropriated by the incumbent equal tax revenues τ:

$$\nu\gamma_{vt} + (1 - \nu)\gamma_{nvt} + r_t = \tau.$$

The priors are that there is a proportion p_b of benevolent incumbents and a proportion $1 - p_b$ of non-benevolent ones. In the second period, there are no reputational concerns. Hence, each type picks its preferred level of public goods provision: benevolent types pick $\left(\gamma^b_{vt+1}, \gamma^b_{nvt+1}\right) = (\tau, \tau)$, while non-benevolent types pick instead $\left(\gamma^{nb}_{vt+1}, \gamma^{nb}_{nvt+1}\right) = \left(\frac{1}{\alpha}, \frac{1}{\alpha}\right)$.[11]

In the first period, voters will want to reelect a benevolent incumbent and replace a non-benevolent one. This introduces reputational concerns in the model since a good reputation is important for getting reelected. Though a benevolent government has no problem in announcing what it actually plans to do because it has nothing to hide, a non-benevolent government must be willing to lie if it diverts resources from the public treasure to its own pockets. Since cheap talk is never informative, only the actual provision of visible public goods counts.

The equilibrium can be either pooling or separating. With a pooling signal, the non-benevolent type mimics the provision of visible public goods by the benevolent type, $\gamma_{vt} = \tau$, and rents are extracted from the under-provision of non-visible public goods, as discussed in the Appendix. The median voter will be indifferent between the incumbent and the challenger because the expected utility in the second period is the same with either candidate: there will be a proportion β of benevolent incumbents and a proportion $1 - \beta$ of non-benevolent ones. Hence, the probability of

[11] With log utility, the marginal utility of private rents for non-benevolent incumbents is α. The level of public goods g_{st+1} that provides that marginal utility, for $s = v, nv$, is $\frac{1}{\alpha} = u^{-1}_{g_{st+1}}(\alpha)$. Above the threshold $\underline{\alpha} = \frac{1}{\tau}$, the higher α is, the lower the provision of public goods; at or below that threshold, rents are null.

reelection $P(\gamma_{vt}^b) \in [0,1]$. Let $P(\gamma_{vt}^b) = \frac{1}{2}$ when the median voter is indifferent.

There might also be an equilibrium with a separating signal $\gamma_{vt} > \tau$ in which the benevolent incumbent provides an extraordinary amount of visible public goods.[12] For a benevolent government, the drawback of a separating signal is that it has a welfare cost because it distorts the optimal provision of public goods.[13] This is where transparency initiatives come in: they may be a less costly way for benevolent governments to signal their type.

Transparency Initiatives

What happens if the government can launch a transparency initiative to reveal part of its plans and their future fulfillment through its post-electoral promises? This characterizes the scenario faced by individuals assigned to the treatment group in our experimental design. We adopt a specification by which transparency initiatives are limited to the interval $\omega \in [0, \nu + \lambda]$: revealing information is not costly until variety $\nu + \lambda$, where $\lambda > 0$, but for $\omega \in (\nu + \lambda, 1]$ there is a prohibitively high information cost. Our motivation is that some things are easier to communicate than others, e.g., certain maintenance costs are hard to report.

We formalize the transparency initiative as a signal that allows honesty to have a bite because it has differential costs for benevolent and non-benevolent governments: it imposes no costs if incumbents are honest, $j = h$; if they are dishonest, $j = nh$, while a proportion $q_{nh,low}$ face no costs, a proportion $q_{nh,high} = 1 - q_{nh,low}$ face a cover-up cost of $K + \kappa(\lambda - \nu)$. This turns the transparency reform into costly talk for dishonest types with high cover-up costs. The budget constraint is now:

$$\nu\gamma_{vt} + (1-\nu)\gamma_{nvt} + r_t - \mathrm{I}(j)(K + \kappa(\lambda - \nu)) = \tau, \quad \text{where} \quad \mathrm{I}(nh, high) = 1, \mathrm{I}(h) = \mathrm{I}(nh, low) = 0.$$

In the second period, there are no reputational concerns. Hence, as before, benevolent types pick $\left(\gamma_{vt+1}^b, \gamma_{nvt+1}^b\right) = (\tau, \tau)$, while non-benevolent types pick $\left(\gamma_{vt+1}^{nb}, \gamma_{nvt+1}^{nb}\right) = \left(\frac{1}{\alpha}, \frac{1}{\alpha}\right)$.

In the first period, a benevolent incumbent assigns the full budget to the provision of public goods, $\left(\gamma_{vt+1}^b, \gamma_{nvt+1}^b\right) = (\tau, \tau)$, and will also launch the reform because it can raise its reputation of

[12] A separating equilibrium may not exist if future reputation is sufficiently important.

[13] It is common across the world for the allocation of public monies to new construction (visible) but much less to maintenance works (non-visible).

trustworthiness (this brings a higher probability of reelection). The choice of non-benevolent types depends on their cover-up costs if they act dishonestly (see Appendix).

When a reform Π is carried out, the posteriors are by Bayes' law:

$$\mu(b|\Pi) = \frac{p_b}{p_b + (1 - q_{nh,high})(1 - p_b)} > p_b \quad \text{if } q_{nh,high} > 0. \tag{1}$$

The polar cases are $q_{nh,high} = 0$, when there is no updating of beliefs, and $q_{nh,high} = 1$, when the reform is a separating signal. In the intermediate cases where the costly talk parameter $0 < q_{nh,high} < 1$, there will be a semi-separating equilibrium where types $j = h$ and $j = nh, low$ launch a transparency reform and provide a high level of visible public goods γ_{vt}^{b}, while type $j = nh, high$ abstains and provides a low level of visible public goods γ_{vt}^{nb}.

2.2 Priors

Our experiment allows us to control for the endogeneity of reputation by discriminating between initial trust (the priors) and final trust (the posteriors). Equation 1 shows there is a circular relationship between transparency initiatives and trust in government because initial trust affects how people process this verbal information. We expect our experiment to affect reputation because it takes place in a relatively high-trust environment.

Since the common prior assumption does not hold, we incorporate heterogeneous priors when calibrating the model, allowing p_b to take the values p_{b1}, p_{b2},..., p_{b7}. Heterogeneous priors imply by equation 1 that new information will not be evaluated the same way by all citizens. In a political context, the heterogeneous priors, which the model represents with the perceptual errors σ_i, can reflect partisan and ideological differences. Citizens may use affective mechanisms to interpret newly acquired information, analyzing the facts through the lens of their party affiliation or their affection toward certain ideology (Slothuus and De Vreese, 2010). Allowing subjective beliefs to differ among individuals is a way of incorporating the effects of politically-motivated reasoning, an issue extensively discussed in the literature.[14]

[14]Regarding this, Gerber and Green (1999) distinguish between evaluations (which reflect preferences regarding presidential policy orientations) and beliefs. They then try to explain the differential presidential approval ratings among Democrats, Republicans and Independents as mainly reflecting differences in evaluations. Bartels (2002), however, shows that politically-motivated reasoning, rather than unbiased learning, is required to explain differences in how beliefs react to the same facts: in comparison to independent voters, partisans over-react to favorable information and under-react to unfavorable information, so he identifies a partisan bias in presidential approval ratings.

Focusing on how to isolate politically-motivated reasoning, Thaler (2022) discovers that this happens with new information because voters tend to trust more information that they find more to their liking. Motivated reasoning not only has an effect on learning but also on priors: due to the biased updating of information in the past, pro-government individuals will have higher trust priors than independent voters, while anti-government individuals will have lower priors.

Allowing for heterogeneous priors makes room for part of the effects of politically-motivated reasoning on Bayesian updating, but there are further effects that we do not control for: the parameter $q_{nh,high}$ in equation 1 has a uniform value in our calibration, so we do not incorporate the potential under-reaction of respondents with low trust (which presumably includes more opposition voters), and over-reaction of respondents with high trust (which presumably includes more government supporters), to good news.[15] Nevertheless, in our data what is salient is something else: heterogeneous learning behaviors. We turn to this now.

2.3 Social and Individual Learning

Since the transparency initiative was launched in 2017, the respondents were asked whether they had heard about the post-electoral promises and were classified into three groups: familiar, somewhat familiar, and unfamiliar.[16]

We did not expect the group "Familiar" to react to the treatments because the provided information is already in their information set. That is indeed what we find. Surprisingly, the treatment only leads the group "Somewhat familiar" to update their priors, closing the gap in trust with the group "Familiar." The "Unfamiliar" group, which could potentially learn the most, does not react to any treatment; rather than sheer ignorance, this suggests that this group seems not to care about what the government says. A possible interpretation of the effect for this group is that these individuals acquire knowledge through their direct observation rather than third-party information.

Schelling (1960) points out on p. 117 that "moves ... have an information content, or *evidence* content, of a different character from that of speech. Talk can be cheap when moves are not." Though we view the transparency initiative as costly talk, rather than cheap talk, both are forms

[15] Our survey experiment lacks information on respondents' partisanship and ideological preferences, so it is not possible to determine how they affect the initial level of trust in government.

[16] The familiarity level is determined by the answer to the question "Are you familiar with the Buenos Aires Elige program?" Individuals could respond "I am familiar with it," "I have heard about it" (that we treat as somewhat familiar), or "I am unfamiliar with it."

of social information, i.e., "things that an individual can learn from others, be it through intentional communication, demonstrations, or the mere observation of behaviors that are not necessarily meant to be seen" (Morin et al., 2021). This contrasts with individual information collected by us directly, which is not open to the noise or deliberate deception of social information (Morin et al., 2021).

Since the transparency initiative is a form of intentional communication, individuals may vary in the extent to which they trust it beyond reasons that have to do with politically motivated reasoning discussed above. For example, Hertz et al. (2021), in an experiment on learning about which of two lakes is best for fishing, find a bimodal distribution that is in line with earlier studies of social learning: some participants fully exploited social information (either the observation of what an expert chose or explicit advice on which of two lakes to fish, where the information was accurate in both cases), while others stuck to a trial-and-error strategy, learning individually from their own fishing experience. Morin et al. (2021) discuss similar two-armed bandit problems where some participants completely ignored the information about other people's choices.

This learning heterogeneity in the population, with some respondents that completely ignore social information, is relevant to our experiment because of the group "Unfamiliar."[17] We also find some differences between the two groups receptive to social information, because the group "Somewhat familiar" trusts government information less than the group "Familiar." We treat these three groups as different types of social learners, so the calibration of the model using equation 1 is adapted in Section 4 to distinguish between different types of information updating.

3 Survey Experiment

We study an actual transparency initiative in Buenos Aires City to assess the value of information about post-electoral promises on government reputation. According to our model, introducing the transparency initiative makes sense if a benevolent incumbent can provide a signal a non-benevolent cannot.

The city has been steadily increasing transparency over the last two decades (Alessandro et al., 2021). Complementing the city's ongoing efforts, the mayor of Buenos Aires made a series of

[17] Since the group "Unfamiliar" is not responsive to the transparency initiative, it may be less effective as a signal than the provision of visible public goods. However, as shown below, the median respondent of our sample seems to be affected by the initiative, so it may still be very effective in terms of the median voter.

promises to residents upon taking office as part of his transparency promotion strategy. These promises are specific and quantifiable objectives that span government sectors and are based on citizens' interests and the United Nations' Sustainable Development Goals. The goals are outlined, and progress toward compliance is reported on the local government's website.[18]

We designed an online questionnaire to elicit information about trust in the government. The questions attempt to capture individuals' perceptions about the competence, benevolence, and honesty of the government, following Grimmelikhuijsen (2012), and are available in Supplementary Material (SM), part B.[19] A total of 2,375 complete interviews were carried out in December 2019 by a company that specializes in collecting online survey data.[20] The sample was stratified with quotas by gender, age group (18 to 60 years old), and socioeconomic status. Within strata, individuals were assigned at random to one of four possible treatment categories: three informational vignettes and a control group. Treated individuals answered the battery of questions on trust in the government after receiving the informational pieces. A description of the timeline of the survey experiment is presented in Figure A1 (SM).

We use a simple design, as recommended by Bouwman and Grimmelikhuijsen (2016), to evaluate the importance of information regarding promises and their impact on trust. Each treatment presents information about the promises made by the government. Treatment 1 mentions the existence and relevance of mayoral promises, providing four examples at the city level but without any details on the level of compliance (SM, Figure B1). Treatment 2 provides the same information as Treatment 1 and shows the government's performance in fulfilling the promises at the city (aggregate) level (SM, Figure B2). Treatment 3 provides the same general information about the promises, but it presents a map with dots for all the achievements at the *comuna* level, highlighting three specific examples of promises fulfilled at that level (SM, Figure B3).[21] It is important to note that individuals in Treatment 3 received an infographic designed for the specific *comuna* they report living in.[22] Unfortunately, by the nature of the implementation of the experiment, there

[18] See `https://buenosaires.gob.ar/compromisos`.

[19] The questionnaire also includes questions about individuals' confidence in politicians and public servants, following Keefer et al. (2018, 2020), which are discussed in SM, part **??**. The full questionnaire in Spanish is available upon request. See Alessandro et al. (2021) for additional details.

[20] For details about the company, visit `http://www.isonomia.com.ar/en/`.

[21] The City of Buenos Aires, Argentina's capital and most populous city, is subdivided into 15 *comunas* that work as territorial, administrative and political entities, and include 48 neighborhoods (SM, Figure B4).

[22] Individuals were asked in the characterization module about the *comuna* they live in; then, the survey program selected the infographic that matched it.

are differences between Treatments 2 and 3 beyond the fact that the first provides examples of fulfillment at the city level and the latter at the *comuna* level. On one hand, Treatment 2 displays four promises, whereas Treatment 3 displays three. On the other, the type of promises is also different and does not necessarily align with the policy area disclosed in Treatment 2 (infrastructure investments, mainly).

4 Data and Empirical Analysis

4.1 Data Description

Table 1 presents descriptive statistics for the main observable characteristics of the respondents and balance on covariates measured before presenting the experimental vignette to participants. The first column shows the sample average and the standard deviation for the control group. The average respondent is female (57%), completed secondary education (nearly 85% of surveyed individuals have completed high school), and is employed (52%).[23] Despite the existence of a dedicated website and the public announcement of post-electoral promises, only 22% of the sample were familiar with them before the intervention took place, while another 42% found them vaguely familiar (SM, Figure A7).

The groups are well balanced; only 4 out of 45 differences are statistically significant at the 10% level, consistent with chance. Based on the balance on observable characteristics, we consider the randomization successful.[24] Additionally, p-values of tests of equality of coefficients identify no systematic differences in sample composition across treatment groups. There are minor disparities between people assigned to the first and second treatment groups regarding gender and educational attainment, with more women in the first treatment arm and more educated individuals in the second.

Dependent variables on reputation are of two types: a direct measure of trustworthiness and

[23] The sample is not necessarily representative of the overall population of the city. In particular, it is slightly more educated, older (the average age in city is about 40 years), and has a higher share of women than the city's population—which is about 53% according to 2010 census data.

[24] We estimate an OLS regression with clustered errors at the *comuna* level to assess balance across treatment assignments. In this specification, observable characteristics act as dependent variables, and treatment variables as independent. To assess balance, we also conducted randomization inference procedures over the difference in means between treated and control units. Adjusted p-values from a thousand Monte Carlo simulations of the OLS regressions provide similar conclusions.

an index that captures various dimensions of the citizenry's perception about the government (SM, Table A1). We have a question on the perception of trustworthiness, which we use as a direct measure of trust. We also work with the individual responses to a series of 11 questions that inquired the participant to show her position on different statements about the Government of the City of Buenos Aires, ranging from *Strongly disagree* to *Strongly agree*. Those questions attempt to identify how the respondent feels about the competence (is capable? does what is best for the city? is innovative? thinks in the long term? plans and informs its plans?), benevolence (acts in the interests of residents? helps those in need? pursues policies my family cares about?), and honesty (is sincere? is transparent? fulfills its promises?) of the city government.

To reduce the dimensionality of the information provided by the 11 questions, we construct summary indexes.[25] To construct the indexes, we exploit a principal component analysis methodology (PCA henceforth) in which the first component explains at least 80 percent of the variance regardless of the index (SM, Table A2).[26] We build three sub-indexes (Competence, Benevolence, and Honesty) and one global index that summarizes all questions on perceptions.[27]

Citizens have relatively high trust in the city government to begin with. Figure 1 provides a first glimpse at the perceptions of the respondents. It looks at the control group's responses to each aspect of government competence, benevolence, and honesty. In general, there is a positive assessment of the government. Respondents consider the government to be capable and innovative, among the top attributes. However, they grade the government lower regarding how much it helps those in need and pursue programs that benefit families.

4.2 Empirical Strategy

We first estimate the following model to understand the effect of providing information about government promises and their fulfillment on government trust:

[25] The aggregation improves statistical power to detect consistent effects across specific outcomes when these specific outcomes also have idiosyncratic variation.

[26] We also applied a factor analysis methodology and consistently found that the first factor explains a significant proportion of the variation. Furthermore, in a parallel analysis, we see that we should stay with a single factor under a decision rule of thumb of one. However, we report results for indices constructed for each dimension of trust, namely, competence, benevolence, and honesty, for interpretation purposes.

[27] Robustness exercises include running the regressions with the individual questions. Conclusions on statistical inference remain the same when we correct p-values using the Westfall and Young procedure (Jones et al., 2019).

$$Y_{ic} = \alpha + \beta_1 T_i^1 + \beta_2 T_i^2 + \beta_3 T_i^3 + \lambda X_{ic} + \epsilon_c, \tag{2}$$

where T^n is the treatment assignment, $n = 1, 2, 3$, depending on the treatment individual i from *comuna* c was exposed to. The treatment arms are as follows: (1) Promises at the city level, (2) Promises and their fulfillment at the city level, and (3) Promises and their fulfillment at the *comuna* level. A respondent was exposed to one treatment arm only, and individuals in each treatment arm were compared against those who did not receive any information.[28] X_{ic} is a vector of controls that includes all observable characteristics available from the survey: age, gender, education (completed high school or college), labor status (employed or unemployed), socioeconomic level (ABC1 describes the group with the highest income), revealed preferences for public budget allocation between education and infrastructure, and pre-treatment beliefs on government quality. We also include *comuna* fixed effects for political divisions within the city.

Sixty percent of respondents in the control group consider that the government is transparent. Since most respondents consider that the government is truthful, our expectation is that providing information about post-electoral promises and their fulfillment matters: $\beta_n > 0$, $n = 1, 2, 3$. City residents can, of course, compare the information received with their personal experience. Thus, promising but not complying with those promises is not good policy in equilibrium; as such, promises by themselves already carry a load of information (Alessandro et al., 2021).[29] Since individuals may feel more comfortable in updating their perceptions if they are also shown information on fulfillment, we expect $\beta_2 \geq \beta_1$.

Targeted information could have added value for the recipient ("the government is not only doing what they promise, but they are doing it in my *comuna*").[30] Although showing information on the achievement of the goals at a more local level could be more informative than presenting the promises alone, we do not have a strong prior on the differentiated effect of T^3 compared to T^1. First, T^3 shows fewer promises than those presented in T^1 due to logistical issues in the im-

[28] As shown in the balance table, the randomization process of the treatment was successful.

[29] This project is carried out with political will and in conjunction with the City of Buenos Aires' local government. The government is aware that making promises and reminding citizens of them has a meaningful information load.

[30] There is evidence that individuals update their priors more when the information they receive refers to a closer or more homogeneous group. See, for example, Miranda et al. (2020) for the case of water consumption. Still, that evidence may not travel well because the information is not about individuals in the *comuna* but rather government action that affected the *comuna*.

plementation phase. Second, T^3 does not consistently present promises in policy areas comparable to those shown in treatments 1 and 2. While promises in T^1 and T^2 mainly concentrate on urban mobility and infrastructure issues, the ones presented in T^3 include health and education, in addition to infrastructure projects. Third, some of the displayed vignettes in T^3 contain information that may not be informative for all citizens from those *comunas*. Imagine the case of information about a public school extended schedule; this exclusively affects those individuals with school-aged children who could be potential beneficiaries of such improvement. Our hypothesis is, therefore, that $\beta_3 \geq \beta_1$. We do not have strong priors regarding the relationship between β_3 and β_2.[31]

4.3 Baseline Estimates: Average Effects

We start by evaluating the effect of the different vignettes on reputation using indices that approximate the trust components we attempt to explain: a general composite index of trust in government and three sub-indexes, referred herein as dimensions, that capture its perceived competence, benevolence, and honesty. We also estimate the effect on a direct measure of trustworthiness.

Table 2 presents the compound index results, and each of the dimensions of trust identified by Grimmelikhuijsen (2012), looking at the disaggregated treatment arms level.[32] The effect of Treatment 2 is consistently higher than the remaining two treatment arms. However, we do not observe significant differences between providing performance information or just informing about the promises.[33] Further, when we compare people who received information on performance at the aggregate (city) versus the local (*comuna*) level we do not observe differences in trust perceptions. Results on each dimension of the index—competence (column 4), benevolence (5), and honesty (6)—are very similar to those of the composite index. The last column of Table 2 depicts the results of providing information on the direct measure of trustworthiness. Again, the three treatments

[31] Table A11 shows the results by *comuna*, considering the third treatment arm nature. There is no regular pattern in the relationship between treatments 2 and 3. As expected, the influence of tailored information on individuals' beliefs about the government is systematically stronger than generic information about the promises. However, we do not observe statistically significant changes in treatment allocations (except for *comunas* 5 and 14). Because mixed results within communes could be explained by the fact that families living in different communes may have characteristics that cause them to respond differently to treatment, we offer in Table A12 a selection analysis that evaluates whether socio-demographic traits and ex-ante perceptions of the government are more widespread in particular communes than others.

[32] Controls and *comuna* fixed effects were incorporated to improve estimate precision. The results are presented in a constructive manner, facilitating the assessment of each feature's impact within the model.

[33] We conducted Wald tests of equality of coefficients in each estimation. We did not find statistically different results with any of the informational treatments.

are positive and significant, but not statistically different Differences between the control and the treated groups (pooled) are of about 0.10 standard deviations (SD) for the composite index and its dimensions (SM, Figure A2).

Table 3 shows the results for the components of each of the three dimensions. The same conclusions as in Table 2 prevail. In general, information about *compromisos* and government performance at the city level lead to higher increases in perceptions, with no statistical differences within treatment arms. These results indicate that, in our sample, i) providing information about promises is valuable for increasing trust, i.e., $\beta_1 > 0$ in terms of equation 2; ii) providing information about the fulfillment of those promises seems to add some but little additional information, i.e., $\beta_2 \geq \beta_1$; and iii) providing information at the *comuna* level does not increase trust more than providing information at the city level, i.e., $\beta_2 \geq \beta_3$. We cannot interpret these results, however, given the caveats about the experiment design discussed earlier. Supplementing information about fulfillment either at the city or comuna level perhaps does not add significantly more to trust than simply providing information about the government promises because many of the respondents have already acquired much of that information first-hand by personal experience in the city streets. As analyzed in the theoretical model, besides the government transparency initiative, another signal is at work: government performance.

4.4 Heterogeneous Priors

The respondents do not share common priors, so we analyze the impacts of the transparency initiative taking into account their differing priors. We then use quality perceptions, which go hand in hand with initial levels of trust in government, as a proxy of initial reputation.

Effects by Trust Scale

The initial beliefs of the respondents are very heterogeneous. The Generalized Ordered Logit specification in Table 4 considers all categories of the agreement scale for each component of the trust dimensions and the direct measure of trustworthiness (see also SM, Figures A3 to A5). People move to higher levels of the trust distribution with all the treatments: category 1 (strongly disagree) falls in 32 out of 36 cases, with a significant fall in 28 cases, while category 7 (strongly agree) always rises, significantly so in 29 cases.

As to the intermediate categories, if the treatments push people up from one category to the next, as in the model, then the effects are cumulative and the biggest jumps in trust are led by the changes in the middle part of the distribution. When we single out the highest category in the 35 cases with a significant fall, there is one case in category 2 (disagree), three in category 3 (somewhat disagree), eight in category 4 (neutral), 18 in category 5 (somewhat agree), and five in category 6 (agree). This suggests that greater levels of trust might be led by positive impacts in the middle part of the distribution. We now look into this issue using initial perceptions about quality of government.

Effects by Perceived Government Quality

To try to understand better how the treatment effects may depend on initial trust, we ask how trust differs in the treatment and control groups for different priors on government quality. This takes advantage of the fact that we have information on initial beliefs on the quality of government from all respondents and that perceptions of the quality of government and trust priors are very closely correlated in the control group.[34]

Our sample thinks highly of the city government's quality. People in the control group gave an average rating of 7.2 points on a scale of 1 to 10, with 1 being the lowest perceived quality and 10 being the highest. Considering the sample sizes in each of the initial variable's categories and for interpretation purposes, we have narrowed the spectrum of perception of government quality to three categories to evaluate heterogeneous effects: low, medium, and high quality. Categories were constructed ad hoc to reflect their definition. Those who ranked the government quality between 1 and 3 were classified as low (L), 4 to 7 as intermediate (M), and 8 to 10 as high (H).

People appear to respond differently to the informational treatments across the perceived quality distribution. Figure 2 shows the effects on trust vary with the perceived quality of the government (see SM, Table A6, for further details in a simplified version considering a general treatment condition). Individuals with the lowest assessments of government quality have wide confidence intervals that do not reject the null hypothesis of no effect; however, they show a positive response to information, although slightly lower than those who have an intermediate evaluation of government

[34]Figure A6 in the SM depicts the correlation between the perceived quality of the government and trust. Overall, the correlation is 0.7603.

quality and were highly receptive to information. People with a very high assessment of government quality did not significantly respond to the information provided through the vignettes, which is consistent with a ceiling effect. If the perception of government quality is a good proxy for initial trust, then these results lend support to the interpretation that the largest effects in Table 4 are in the intermediate categories of trust.

4.5 Previous Knowledge of Transparency Initiative

We now explore our framework's prediction that individuals with ex-ante information about post-electoral promises will not respond to the informational treatments, so the effects should be concentrated in the groups 'Unfamiliar," who are completely unaware of the post-electoral promises, and "Somewhat familiar," who are only slightly familiar with the transparency program. We then look at how trust differs among these groups, to see if this can help understand differences in their previous knowledge of the transparency initiative.

Effects by Previous Knowledge

Despite the existence of a dedicated website and its public announcement, the participants vary widely in their familiarity with the transparency initiative: 22% of the participants were familiar (F) with the promises before the survey experiment, 42% were somewhat familiar (S), and 36% were unfamiliar (U).

We observe in Figure 3 that people who previously knew the promises already have high assessments of the city government's competence, benevolence, and honesty; thus, they may have already incorporated this information into their trust perceptions. Figure A8 (SM) shows a positive gradient in the relationship between previous knowledge of the policy and assessments of the city government. Individuals within the control group unfamiliar with the policy, who are unaware of or have never heard of it, start with a lower level of trust in the government and each of its components. This level increases as people acquire information about the promises. Another way of looking at this is using the perceived quality of government as a proxy of initial trust. Figure A9 (SM) depicts the distribution of the perceived quality scores by each level of knowledge. The distribution of people who previously knew about the pledges is skewed right in comparison to those who had heard of it, while those who did not know about the pledges are marginally shifted

to the left.[35]

Previous knowledge of the policy was not randomly assigned, yet from Table 1, we observe that the proportion of people who knew about the policy is not statistically different among treatment statuses (which were randomly assigned). People who receive information about promises and their fulfillment at the city level are marginally more aware of the policy's existence. Considering that treatment arms have roughly the same proportion of people who know the policy ex-ante, and they are assigned at random, we explore how differing levels of previous knowledge of the policy affect the treatments.

Our framework leads us to anticipate that previous knowledge of the post-electoral promises can strongly affect the response to informational treatments. Figure 4 shows the heterogeneous effects of previous knowledge of the governmental promises. As before, we observe that the second treatment arm, T^2, has a slightly higher impact on trust than T^1 and T^3, although the difference is not statistically significant. However, this result depends on the initial information set. People who previously knew about the "Compromisos" policy do not significantly respond to the information provided through the vignettes, which confirms our framework's prediction that those familiar with the pledges would be unaffected by the treatment because their information set remains unchanged. Surprisingly, participants who were exposed to information about the transparency policy for the first time were much less receptive to the vignettes than those who had already heard something about the policy.

Table A7 (SM) provides the regression findings of a simplified version interacting previous knowledge with being treated, regardless of the information received. It shows that the treatment closes the average gap in trust between the groups somewhat familiar and familiar with the pledges. On the contrary, we reject the null hypothesis that the treatment closes that gap in trust perceptions for individuals who are first exposed to the pledges through the survey. Figure 5 additionally shows that the distribution of the intermediate group S is the only one that significantly shifts to the right, though not enough to match the distribution of those already familiar with the initiative.

[35]The two-sample Kolmogorov-Smirnov test for equality of distributions rejects the null hypothesis of no difference between the three distributions.

Heterogeneous Learning Behaviors?

One would expect trusted sources to affect beliefs more than mistrusted sources, because trust is crucial for new information to affect beliefs.[36] This can help explain the differences in the respondents regarding their previous information about the transparency initiative. Figure 6 shows that trust in different sources of information varies strongly among the groups of individuals according to their previous knowledge of the policy.

Group U, which is unfamiliar with the policy, stands out. Figure 6 shows that this group does not trust much any information from others: while it has some trust in the family (though much less than groups F and S), it has very little trust in information from the government, social media, or traditional media. Their distrust of verbal information fits the pattern in Hertz et al. (2021) of people who are individual learners rather than social learners. As mentioned before, the fact that they heard nothing at all might thus really mean something else: that they ignored the pledges because they only believe what they see. This group might thus react to observed performance rather than to transparency initiatives.

On the other hand, groups S and F are remarkably similar in their trust of family, social media, and TV or radio; the only noticeable difference is that group S trusts the government a bit less, though the difference is not statistically significant. If people rely more on the most trusted sources to get information, this might help explain why the sources through which groups F and S gather information about "Compromisos" differ. Group F, the most familiar with the policy, obtained information primarily from official sources: around 60% had visited its website, and an additional 15% were aware of its existence. Group S had heard about the policy mainly through social media and pamphlets.[37]

In summary, the transparency reform does not seem to work equally well with all kinds of respondents. Only the most politically informed fully incorporated it to their information set.[38]

[36] In Thaler (2022), motivated reasoning leads individuals to give more credence than warranted to the information one would want to believe (in his experiment, the contents of the messages). This entails the insight that people update beliefs more when they trust more, something that applies broadly to all learning.

[37] Group F was 9.3 pp more likely than group S to have known them by attending a mayor's presentation, 8.1 pp by visiting the city's website, and 3.6 pp by reading CABA articles or notes, with only a small percentage visiting the policy website (see Table A9).

[38] Piotrowski et al. (2019) discuss the differences between more and less politically informed citizens. While they find that citizens who actively engage with politicians and the political process tend to be more critical when presented with information about the government's achievements, in our sample the group of individuals that is the most politically engaged has the highest trust levels; the fact that they do not react to the treatment is due to another

However, it also seems to have been somewhat effective with the people active on social media. If there is a spillover effect of government information through social media, the reform might end up being quite a potent signal.

4.6 Sorting Out the Effects

To econometrically sort out the impacts of trust priors (proxied by perceived government quality) and previous knowledge of the transparency initiative, we add an interaction term between both variables. We then calibrate our workhorse model to see how much it can explain of the shifts in the trust distributions when we distinguish between heterogeneous learning behaviors.

Interaction Effects

Table A8 (SM) presents a triple interaction identification strategy of treatment T (versus control C), previous knowledge of the transparency initiative F, S, or U, and initial perceptions of government quality H, M, or L on both the trust index and the direct measure of trust. Table 5 summarizes these conditional treatment effects.

The results above on heterogeneous priors only show significant positive effects of the treatment for group M that has an intermediate perception of government quality. When we interact this with the variable previous knowledge, the effects on both the index and the direct measure of trust remain highly significant for group S. However, the effects are not as robust for groups F (only significant when the direct measure of trust is used) and U (only significant when the trust index is used).

As to previous knowledge of the transparency initiative, the results above only show significant effects of the treatment for the intermediate group S. That effect is overwhelmingly confirmed here: what stands out are the consistently significant positive effects whether perceptions of government quality are high, medium, or low.

Insofar as individuals who gave the government a low rating L are more likely to be opposed to it, those who gave it a high rating H are more likely to be its supporters, and those with an intermediate rating M are less likely to be politically radical, we expect perceived government

reason: this information is not new for them.

quality to be affected by political ideology.[39] Given this, motivated reasoning leads to expect the treatment effects to be smallest among individuals with a low rating of government quality L. While the effects are indeed insignificant in subgroups U and F, it is striking that the largest positive effects appear precisely here, among subgroup S. [40]

Calibration

We can derive a point estimate of $q_{nh,high}$, the probability that a non-benevolent government has high cover-up costs and will not launch the administrative reform, using equation 1. Since the final reputation (the posteriors, in the treatment group) is higher than the initial reputation (the priors, in the control group), the equation yields $q_{nh,high} = 0.12 > 0$ with the direct measure of trust. Hence, the transparency initiative leads to a partial updating of beliefs.[41]

This reasoning extends from a setup with homogeneous priors to one with heterogeneous priors. We use the seven categories of the direct measure of trust to calibrate the model. Relying on the assumption that the responses are uniformly distributed within each category, Table 6 shows how well the model can replicate the differences between the treatment and control groups.[42] Picking the value $q_{nh,high} = 0.14$ that maximizes the proportion of squared deviations explained by the model leads to correctly predict, with one exception, that categories 1 to 4 decrease and categories 5 to 7 increase (the miss is that it predicts that category 2 will increase).

The calibration with the whole sample is not our preferred specification because reactions depend on previous knowledge of the transparency initiative. Choosing parameters $q_{nh,high}$ that maximize the fit for each group, the group "Somewhat Familiar" has $q^S_{nh,high} = 0.20$ that correctly

[39] Butler et al. (2021) point out that those with a moderate view of politics are less concerned with politicians' policy positions and more concerned with problem-solving.

[40] On the other hand, though insignificant, individuals in subgroup F of group L show a large negative effect of the treatment; this suggests that they might not only fail to update priors in response to positive information, but also get upset when reminded of something they already dismissed. Related to this, when individuals receive (accurate) information contradicting their beliefs from sources they do not trust, they may interpret it as an attempt to manipulate them, reinforcing their initial perceptions (Aruguete et al., 2021; Keefer et al., 2022).

[41] If $q_{nh,high} = 0$, the equilibrium would be pooling and the treatment group would have the same beliefs as the control group; if $q_{nh,high}$=1, the equilibrium would be separating and the treatment group would fully trust the government. Instead, $q_{nh,high} = 0.12$ corresponds to a semi-separating equilibrium. When we distinguish previous knowledge, $q^S_{nh,high} = 0.20 > 0$, while $q^U_{nh,high} = 0.03$ and $q^F_{nh,high} = 0.02$ are practically nil.

[42] We divide the unit interval into seven categories of equal length: $[0, .143]$, $[.143, .286]$, $[.286, .429]$, $[.429, .571]$, $[.571, .714]$, $[.714, .857]$, $[.857, 1]$. We then invert equation 1 to find the priors that correspond to the posteriors that separate trust categories $i = 1, 2, ..., 7$: $p_{bi} = \frac{\mu_i(1-q_{nh,high})}{1-\mu_i q_{nh,high}}$. This allows us to see what percentage of the control group shifts to a higher category with the treatment according to the model. Except for respondents with either no trust at all or full trust in the government, the model predicts an updating of beliefs.

predicts that categories 1 to 4 fall, and categories 5 to 7 rise. Unlike the group "Somewhat Familiar," where the treatment effects explain a sizable proportion of the shift, the groups "Familiar" and "Unfamiliar" have $q^F_{nh,high} = 0.10$ and $q^U_{nh,high} = 0.11$ that explain next-to-nothing. This agrees with Figure 5 above, where the trust distribution of the intermediate group S is the only one that significantly shifts to the right.

A final caveat should be kept in mind. This calibration does not take into account the fact that the groups "Familiar" and "Somewhat Familiar" incorporated information about the transparency initiative on their own. Hence, the parameters in Table 6 give a lower bound on the effects of the transparency initiative. Though the group "Unfamiliar," which has the lowest trust levels, may differ from the other two groups for a variety of reasons, in Table 7 we use the "Unfamiliar" control group to estimate a very speculative upper bound for the full effects of the initiative, experimental and non-experimental: $q^F_{nh,high} = 0.77 > q^S_{nh,high} = 0.61 > q^U_{nh,high} = 0.11$. This is the ordering we expected, given the differences of trust in government of the three groups. While the median belief in the trustworthiness of the city government is 4 both for the control and the treatment subgroups unfamiliar with the initiative, Table 6 shows that in the whole sample the median is 5 both for the control and treatment subgroups. This very speculative upper bound suggests that the transparency initiative might have pushed the median evaluation in the whole sample from 4 to 5, i.e., from neutral to above average, thanks to the groups somewhat or very familiar with the initiative.

5 Conclusions

Reputation and initiatives to promote transparency are endogenous variables, so it is difficult to disentangle their causal relationships. To evaluate the conditions that make initiatives more likely to succeed, we developed a theoretical model and conducted an empirical experiment in the City of Buenos Aires. The transparency initiative entailed revealing the plans of the local administration as well as its achievements.

The model shows that transparency initiatives allow benevolent governments to signal their type without having to resort to costly misallocations of public goods so common across the world, with too much spending on visible compared to non-visible goods. Non-benevolent governments

will be less likely to introduce them as long as hiding their cheating is costly.

Our framework suggests that the effect of transparency initiatives is critically dependent on priors, i.e., baseline trust levels, because people at the extremes of no trust and full trust will not be affected by the initiatives. Furthermore, the framework predicts that the treatment will not affect those familiar with the policy, because they are not receiving new information. The experimental results corroborate both predictions, showing that the impact of the policy is greatest for intermediate levels of trust, and for those vaguely familiar with the transparency policy. Surprisingly, the experimental results go beyond our initial expectations. The data show a stark difference between people who did not know anything at all about the promises and those who had heard something about them, since all the treatments' effects are through the latter group, which was vaguely familiar.

While the "Unfamiliar" group had, on average, lower initial trust levels than the "Somewhat familiar" group, this does not explain the difference, because even individuals with higher initial trust in the former group react less to the treatment. This result might have to do with heterogeneous learning behaviors: some people only trust what they observe directly, not what others say. Unlike public good provision, which has to do with the direct experience of citizens, a transparency initiative is a very different kind of signal: it is a verbal signal that is not cheap talk only if there are cover-up costs for a substantial proportion of dishonest types. The experiment thus suggests that transparency initiatives may be useful as a complementary tool to solid performance and efficient administrative management.

This empirical finding feeds back into the signaling model, which can accommodate these features by differentiating among individuals according to their learning modes. As there is a group of people unaware of the pledges that seem to be affected mainly by deeds, not words, transparency reforms may be less effective in changing citizens' perceptions than performance. However, the signal might still be effective in terms of the median voter if these groups are not too large, as in our sample.

Another finding is that the treatment effects on people who had only heard of the program, but were not well familiarized with it, are of the same magnitude as the initial gap in trust between people aware of the commitments and people who had only heard of them: the effects are between 0.14 and 0.19 SD. This suggests that the impact of the vignettes is not only short-term. Hence,

unlike Marvel (2016), the effects of the treatment (here, post-electoral promises) seem to be long-lasting, perhaps because they are backed by government performance that is seen as good by most respondents.

The policy implications are straightforward. First, since overall transparency makes it harder for non-benevolent actors to hide their actions, transparency initiatives provide benevolent governments the opportunity to signal their type without having to incur in costly misallocations of resources. Hence, there are conditions under which it is incentive-compatible for politicians to introduce transparency initiatives. Second, what matters for the success of the initiative is not only average reputation, but also how reputation is distributed among the population. Transparency initiatives will be less effective in a very polarized society than in a society with independent citizens interested in actual government performance. Finally, transparency initiatives complement the direct experience that citizens have with government, which may help explain why all informational treatments elicited a positive reaction from respondents, in a setting where a great deal of citizens already have substantial initial trust in government and perceive, on average, that government quality is high. These results should encourage governments to follow policies of making and keeping promises, ensuring good government performance.

This article evaluates the conditions under which transparency initiatives are more likely to emerge, the role of beliefs and the process of information gathering, and a rationale behind the allocation between visible and non-visible public goods. Still, some questions remain. On the one hand, the promises were chosen by the government. On the other, we cannot evaluate if the commitments (and the evolution of compliance) affected other public policy objectives. As such, we cannot shed light on whether, in terms of the overall portfolio of public policies, promises before execution are better or if "the intention to take action should not be revealed before the consequences of the action are observed" (Prat, 2005, 869) because it affects effort allocation across tasks. However, this article can be a solid foundation for others to build on.

References

Alessandro, M., Cardinale Lagomarsino, B., Scartascini, C., Streb, J., and Torrealday, J. (2021). Transparency and trust in government. evidence from a survey experiment. *World Development*, 138.

Ansolabehere, S. and Snyder, J. M. J. (2000). Valence politics and equilibrium in spatial election models. *Public Choice*, 103:327–336.

Ardanaz, M., Otálvaro-Ramírez, S., and Scartascini, C. (2023). Does citizen participation in budget allocation pay? A survey experiment on political trust and participatory governance. *World Development*, 163.

Aruguete, N., Calvo, E., and Ventura, T. (2021). News by popular demand: Ideology, reputation, and issue attention in social media news sharing. *International Journal of Press/Politics*.

Ashworth, S. (2012). Electoral accountability: recent theoretical and empirical work. *Annual Review of Political Science*, 15:183–201.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24(2):117–150.

Besley, T. and Coate, S. (1997). An Economic Model of Representative Democracy. *The Quarterly Journal of Economics*, 112(1):85–114.

Blanco, L. and Ruiz, I. (2013). The impact of crime and insecurity on trust in democracy and institutions. *American Economic Review*, 103(3):284–88.

Bouwman, R. and Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management*, 29(2):110–131.

Butler, D. M., Hughes, A. C., Volden, C., and Wiseman, A. E. (2021). Do constituents know (or care) about the lawmaking effectiveness of their representatives? *Political Science Research and Methods*, pages 1–10.

Evans, D. K., Holtemeyer, B., and Kosec, K. (2019). Cash transfers increase trust in local government. *World Development*, 114:138 – 155.

Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–745.

Gerber, A. and Green, D. (1999). Misperceptions about perceptual bias. *Annual Review of Political Science*, 2(1):189–210.

Grimmelikhuijsen, S. (2012). Linking transparency, knowledge and citizen trust in government: an experiment. *International Review of Administrative Sciences*, 78(1):50–73.

Hamm, J. A., Smidt, C., and Mayer, R. C. (2019). Understanding the psychological nature and mechanism of political trust. *PLoS ONE*, 14(5):e0215835.

Hertz, U., Bell, V., and Raihani, N. (2021). Trusting and learning from others: immediate and long-term effects of learning from observation and advice. *Proceedings Royal Society B*, 288.

Jones, D., Molitor, D., and Reif, J. (2019). What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study*. *The Quarterly Journal of Economics*, 134(4):1747–1791.

Keefer, P., Scartascini, C., and Vlaicu, R. (2018). Shortchanging the Future: The Short-Term Bias of Politics. In Izquierdo, A. and Pessino, C., editors, *Better Spending for Better Lives*, chapter 10. Inter-American Development Bank, Washington, DC, USA, 1 edition.

Keefer, P., Scartascini, C., and Vlaicu, R. (2020). Voter Preferences, Electoral Promises and the Composition of Public Spending.

Keefer, P., Scartascini, C., and Vlaicu, R. (2022). Demand-side determinants of public spending allocations: Voter trust, risk and time preferences. *Journal of Public Economics*, 206:104579.

Kendall, C., Nannicini, T., and Trebbi, F. (2015). How do voters respond to information? evidence from a randomized campaign. *American Economic Review*, 105(1):322–53.

Khemani, S., Dal Bo, E., Ferraz, C., Finan, F., Stephenson, C., Odugbemi, A. M., Thapa, D., and Abrahams, S. D. (2016). Making politics work for development: Harnessing transparency and citizen engagement. Technical report. World Bank Policy Research Report.

Marvel, J. D. (2016). Unconscious bias in citizens' evaluations of public sector performance. *Journal of Public Administration Research and Theory*, 26(1):143–158.

McFadden, D. (1981). Economic Models of Probabilistic Choice. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Economic Applications*. MIT Press, Cambridge, MA.

Miranda, J. J., Datta, S., and Zoratto, L. (2020). Saving Water with a Nudge (or Two): Evidence from Costa Rica on the Effectiveness and Limits of Low-Cost Behavioral Interventions on Water Use. *The World Bank Economic Review*, 34(2):444—463.

Morin, O., Jacquet, P. O., Vaesen, K., and Acerbi, A. (2021). Social information use and social information waste. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376.

Piotrowski, S., Grimmelikhuijsen, S., and Deat, F. (2019). Numbers of narratives? how government message strategies affect citizens' attitudes. *Public Performance and Management Review*, 42(5):1005–1028.

Prat, A. (2005). The wrong kind of transparency. *American Economic Review*, 95(3):862–877.

Rogoff, K. and Sibert, A. (1988). Elections and macroeconomic policy cycles. *The review of economic studies*, 55(1):1–16.

Sances, M. W. (2021). Presidential approval and the inherited economy. *American Journal of Political Science*.

Schelling, T. (1960). *The Strategy of Conflict*. Harvard University Press.

Schoorman, F. D., Mayer, R. C., and Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future.

Slothuus, R. and De Vreese, C. (2010). Political parties, motivated reasoning and issue framing effects. *Journal of Politics*, 72(3):630–645.

Thaler, M. (2022). The fake news effect: Experimentally identifying motivated reasoning using trust in news.

Woodford, M. (2020). Modeling Imprecision in Perception, Valuation and Choice. *Annual Review of Economics*, 12:579–601.

Appendix

Pooling Equilibrium

The text first considers what happens if visible public goods are the only available signal. Let the pooling signal be $\gamma_{vt} = \tau$, the optimal signal for the benevolent type. Then $\left(\gamma_{vt}^{b}, \gamma_{nvt}^{b}\right) = (\tau, \tau)$, $\left(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb}\right) = \left(\tau, \frac{1}{\alpha}\right)$. Plugging the optimal $t+1$ solutions derived above in the utility function $U(.)$, in a pooling equilibrium the expected indirect utility of the non-benevolent type has to be larger if it mimics the benevolent type:

$$\mathbb{E}\left[V^{nb}(\gamma_{vt}^{b}, \gamma_{nvt}^{nb})\right] > \mathbb{E}\left[V^{nb}(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb})\right]. \tag{3}$$

If the non-benevolent type mimics the benevolent type in the first period, it will have a positive probability of reelection $P(\gamma_{vt}^{b})$:

$$\begin{aligned}\mathbb{E}\left[V^{nb}(\gamma_{vt}^{b}, \gamma_{nvt}^{nb})\right] =& u(\gamma_{vt}^{b}, \gamma_{nvt}^{nb}) + \alpha(\tau - \upsilon\gamma_{vt}^{b} - (1-\upsilon)\gamma_{nvt}^{nb}) \\ &+ \delta P(\gamma_{vt}^{b})\left[u(\gamma_{vt+1}^{nb}, \gamma_{nvt+1}^{nb}) + \alpha(\tau - \upsilon\gamma_{vt+1}^{nb} - (1-\upsilon)\gamma_{nvt+1}^{nb})\right] \\ &+ \delta(1 - P(\gamma_{vt}^{b}))\left[\beta u(\gamma_{vt+1}^{b}, \gamma_{nvt+1}^{b}) + (1-\beta)u(\gamma_{vt+1}^{nb}, \gamma_{nvt+1}^{nb})\right].\end{aligned}$$

If it instead separates out, $P(\gamma_{vt}^{b}) = 0$:

$$\begin{aligned}\mathbb{E}\left[V^{nb}(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb})\right] =& u(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb}) + \alpha(\tau - \upsilon\gamma_{vt}^{nb} - (1-\upsilon)\gamma_{nvt}^{nb}) \\ &+ \delta\left[\beta u(\gamma_{vt+1}^{b}, \gamma_{nvt+1}^{b}) + (1-\beta)u(\gamma_{vt+1}^{nb}, \gamma_{nvt+1}^{nb})\right].\end{aligned}$$

Rearranging terms, a non-benevolent incumbent resigns some utility from rents today (net of utility gain from more visible public good provision now) if this opens the door to even more utility from rents in the future (net of utility loss from lower future provision of both types of public goods):

$$\begin{aligned}\alpha\upsilon(\gamma_{vt}^{b} - \gamma_{vt}^{nb}) - \upsilon\left[u(\gamma_{vt}^{b}) - u(\gamma_{vt}^{nb})\right] <& \delta P(\gamma_{vt}^{b})\alpha\left(\tau - \upsilon\gamma_{vt+1}^{nb} - (1-\upsilon)\gamma_{nvt+1}^{nb}\right) \\ &- \delta P(\gamma_{vt}^{b})\beta\left[u(\gamma_{vt+1}^{b}, \gamma_{nvt+1}^{b}) - u(\gamma_{vt+1}^{nb}, .\gamma_{nvt+1}^{nb})\right].\end{aligned}$$

Semi-separating Equilibrium

We here consider what happens when the benevolent type can launch a transparency reform, to supplement the information provided by visible public goods. If a non-benevolent type nb mimics the provision of visible public goods undertaken by a benevolent type, the benefits of a reform for types who act dishonestly and have type $l = nh, low$ are not affected, but types $l = nh, high$ are negatively affected by the cover-up costs of the rents extracted from the underprovision of public goods in the first period:

$$B^{nb,high} = -K - \kappa(\lambda - \nu) < 0.$$

The reform can work as a semi-separating signal if it is only adopted by benevolent types and by non-benevolent types that face no cover-up costs. Using the indirect utility function $V(.)$, the expected utility for the non-benevolent type that faces cover-up costs by mimicking the visible expenditure of benevolent types and adopting the reform, Π has to be smaller than if it does not, $\sim \Pi$,

$$\mathbb{E}\left[V^{nb,high}\left(\gamma_{vt}^{b}, \gamma_{nvt}^{nb}, \Pi\right)\right] \leq \mathbb{E}\left[V^{nb,high}\left(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb}, \sim \Pi\right)\right]. \tag{4}$$

Combining the results from the signal γ_{vt}^{b} with the additional term that captures the impact of the reform, these conditions imply that

$$\mathbb{E}\left[V^{nb,high}\left(\gamma_{vt}^{b}, \gamma_{nvt}^{nb}, \Pi\right)\right] - \mathbb{E}\left[V^{nb,high}\left(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb}, \sim \Pi\right)\right] = \mathbb{E}\left[V^{nb}\left(\gamma_{vt}^{b}, \gamma_{nvt}^{nb}\right)\right] - \mathbb{E}\left[V^{nb}\left(\gamma_{vt}^{nb}, \gamma_{nvt}^{nb}\right)\right] + B^{nb,high} \leq 0.$$

Tables and Figures

Table 1: Summary statistics and Randomization Balance

	Control	Diff wrt. Control			p-value Wald test equality coefficients				Sample
Variable	(av. & s.d.)	T1	T2	T3	T1=T2=T3	T1=T2	T1=T3	T2=T3	Size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gender	0.570	0.042	-0.011	0.005	0.197	0.084	0.168	0.538	2,375
	(0.021)	(0.036)	(0.024)	(0.033)					
Age	47.656	-1.829	-2.155**	-2.146*	0.942	0.740	0.780	0.992	2,375
	(1.399)	(1.165)	(0.956)	(1.099)					
College	0.404	0.007	0.015	-0.011	0.615	0.747	0.426	0.366	2,375
	(0.042)	(0.032)	(0.037)	(0.034)					
High school	0.846	-0.022*	0.025	-0.001	0.047	0.041	0.250	0.395	2,375
	(0.016)	(0.011)	(0.021)	(0.017)					
Employed	0.522	0.039	0.017	0.009	0.630	0.449	0.366	0.765	2,375
	(0.021)	(0.030)	(0.026)	(0.031)					
Unemployed	0.152	-0.018	0.010	-0.008	0.307	0.133	0.557	0.413	2,375
	(0.018)	(0.022)	(0.024)	(0.028)					
Socio-economic level (high)	0.244	-0.025	-0.040**	-0.038	0.656	0.463	0.557	0.938	2,375
	(0.025)	(0.020)	(0.017)	(0.027)					
Voluntary Health Insurance	0.430	-0.007	-0.013	0.011	0.768	0.843	0.584	0.483	2,375
	(0.026)	(0.028)	(0.021)	(0.015)					
Internet at home	0.526	0.000	0.000	-0.013	0.908	0.986	0.666	0.734	2,375
	(0.023)	(0.025)	(0.022)	(0.021)					
Credit Card	0.347	0.016	-0.010	0.006	0.697	0.411	0.687	0.520	2,375
	(0.027)	(0.019)	(0.025)	(0.017)					
One or more cars	0.175	0.005	-0.015	0.001	0.625	0.354	0.877	0.483	2,375
	(0.018)	(0.021)	(0.018)	(0.025)					
Perc. Quality of Governm.	7.200	-0.142	-0.055	0.197	0.088	0.616	0.068	0.064	2,331
	(0.103)	(0.154)	(0.098)	(0.126)					
Knowledge of 'Compromisos'	0.219	-0.001	0.047	0.005	0.289	0.123	0.728	0.205	2,375
	(0.016)	(0.015)	(0.030)	(0.020)					
Trust Others	0.662	0.003	-0.026	-0.022	0.589	0.333	0.396	0.867	2,265
	(0.021)	(0.028)	(0.027)	(0.020)					
Collective Action	0.671	0.010	-0.009	0.012	0.603	0.444	0.936	0.344	2,261
	(0.020)	(0.030)	(0.029)	(0.032)					

Notes: Column (1) shows the sample average and the standard deviation in parentheses for the control group. Columns (2)-(4) show the regression coefficient and the standard error in parentheses corresponding to an OLS regression - observable is the dependent variable and the treatment variables are the independent ones (T1-T3). Columns (5)-(8) show the p-value of a Wald test of equality of coefficients. Finally, column (9) shows the sample size. *Gender* is a indicator variable for women. *Age* is a continuous variable from 18 to 100 years old. *College* takes the value of one when the individual has a college degree at least, and *High school* is read similarly. *Employed* and *Unemployed* are binary variables for those who have full-time employment (or work from/at home) and those who are looking for a job at the time of the survey, respectively. *Socio-economic level (High)* is a binary variable for those with the highest category in socio-economic level. *Perceived Quality of the Government* is self-explanatory and takes values between 1 and 10, in which the lowest value reflects a very bad score while the greatest an excellent score. *Knowledge* is a binary variable and takes the value one if the participant knows the 'Compromisos' policy and zero otherwise. *Trust Others* is a binary variable that takes the value of one when participants indicate that others are reliable or very reliable. *Collective Action* is a dummy variable that indicates whether participants believe that they would be able to collect 500 signatures to support a petition for the government among their neighbors. Clustered errors at the *Comuna* level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Treatment Effect on Trust in Government (by dimension)

VARIABLES	Trust in the Government Global Index			Dimensions of Trust Competence	Dimensions of Trust Benevolence	Dimensions of Trust Honesty	Direct measure Trustworthiness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T1: Commitments	0.051	0.092**	0.091**	0.084**	0.092*	0.092**	0.126***
	(0.053)	(0.032)	(0.032)	(0.032)	(0.043)	(0.034)	(0.031)
T2: Commitments + Fulf. city	0.116**	0.112***	0.112***	0.117***	0.106**	0.097**	0.109**
	(0.046)	(0.032)	(0.033)	(0.029)	(0.045)	(0.034)	(0.037)
T3: Commitments + Fulf. comuna	0.129**	0.097***	0.097***	0.103***	0.113***	0.061*	0.076*
	(0.056)	(0.031)	(0.033)	(0.033)	(0.035)	(0.034)	(0.041)
Constant	-0.070*	-2.417***	-2.375***	-2.335***	-2.120***	-2.397***	-2.335***
	(0.038)	(0.100)	(0.095)	(0.095)	(0.112)	(0.093)	(0.130)
Observations	2,375	2,184	2,184	2,184	2,184	2,184	2,184
R-squared	0.003	0.672	0.674	0.648	0.603	0.634	0.609
Joint significance (p-value)	0.483	0.847	0.853	0.732	0.915	0.592	0.583
T1=T2	0.268	0.636	0.640	0.456	0.773	0.902	0.692
T1=T3	0.309	0.913	0.908	0.719	0.680	0.481	0.321
T2=T3	0.836	0.670	0.678	0.722	0.834	0.354	0.453
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Comuna FE	No	No	Yes	Yes	Yes	Yes	Yes

Notes: All dependent variables are constructed using a PCA method, and standardized with mean zero and standard deviation one. The Competence dimension considers the assessments of the following characteristics, the government: is capable, does what is best for the city, is innovative, thinks in the long run, and plans and informs its plans; the Benevolence dimension considers the following: acts in the interests of its residents, helps those in need and pursues policies and projects beneficial for the families. Finally, the Honesty dimension takes into account: is sincere, is transparent, fulfills its promises. We also study the effect of information on a direct measure of trust in government that asks the respondents to indicate the degree in which she agrees that the city government is trustworthy. Control variables include: age, gender, socio-economic level, labor status, public policy preferences (revealed preferences for public budget allocation between education and infrastructure), being first exposed to information on the experiment, pre-treatment beliefs on government quality and the collective action dummy variable. Clustered errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3: Treatment Effect on Trust in Government (by component)

	Competence					Benevolence			Honesty		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
T1: Commitments	0.069*	0.056*	0.036	0.119***	0.100*	0.044	0.118**	0.095**	0.094**	0.058	0.111***
	(0.039)	(0.027)	(0.039)	(0.038)	(0.049)	(0.044)	(0.043)	(0.042)	(0.036)	(0.039)	(0.026)
T2: Commitments + Fulf. city	0.090**	0.089**	0.053	0.136***	0.162***	0.068*	0.137**	0.090*	0.064	0.057	0.158***
	(0.033)	(0.037)	(0.041)	(0.035)	(0.038)	(0.037)	(0.049)	(0.050)	(0.043)	(0.033)	(0.031)
T3: Commitments + Fulf. comuna	0.102**	0.074*	0.055	0.138***	0.096**	0.054	0.121***	0.141***	0.052	0.016	0.108***
	(0.037)	(0.038)	(0.051)	(0.037)	(0.036)	(0.040)	(0.040)	(0.038)	(0.041)	(0.034)	(0.029)
Constant	-2.176***	-2.069***	-2.166***	-2.172***	-1.894***	-1.964***	-1.960***	-1.994***	-2.165***	-2.257***	-2.438***
	(0.095)	(0.121)	(0.088)	(0.154)	(0.101)	(0.109)	(0.122)	(0.110)	(0.103)	(0.119)	(0.085)
Observations	2,184	2,184	2,184	2,184	2,184	2,184	2,184	2,184	2,184	2,184	2,184
R-squared	0.567	0.602	0.517	0.513	0.446	0.565	0.502	0.508	0.575	0.573	0.589
Joint significance (p-value)	0.844	0.681	0.941	0.914	0.0767	0.824	0.867	0.437	0.689	0.446	0.510
T1=T2	0.672	0.394	0.743	0.686	0.140	0.579	0.728	0.933	0.544	0.994	0.315
T1=T3	0.606	0.713	0.773	0.733	0.947	0.866	0.957	0.365	0.397	0.334	0.939
T2=T3	0.846	0.718	0.960	0.975	0.124	0.750	0.629	0.237	0.764	0.271	0.279

Notes: All regressions include controls and commune fixed effects. All dependent variables are standardized with mean 0 and standard deviation 1. Each column presents the result for perceived performance of the CABA government. The first column displays the global effect on the Index of Trust in the Government. Following Grimmelikhuijsen (2012), next five columns reflect Government Competence: (1) is capable, (2) does what is best for the city, (3) is innovative, (4) thinks in the long-term, and (5) plans and informs; following three columns show Benevolence: (6) acts in the interests of the residents, (7) helps those in need, (8) pursues policies and projects that are beneficial for families; next three, Honesty: (9) is sincere, (10) is transparent, (11) fulfills its promises. Control variables include: age, gender, socio-economic level, labor status, public policy preferences (revealed preferences for public budget allocation between education and infrastructure), first exposure to 'Compromisos' and pre-treatment beliefs on government quality. Clustered errors are shown in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4: Average Marginal Treatment Effect on Trust in Government: Generalized Ordered Logit

Panel A: Competence

	1	2	3	4	5	6	7
The CABA Government...	**Strongly disagree**	**Disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Agree**	**Strongly agree**
Competence 1: is capable							
T1: Commitments	-0.009**	-0.004**	-0.005**	-0.006**	-0.007*	-0.003*	0.033**
	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.016)
T2: Commitments + Fulf. city	-0.010***	-0.004**	-0.006***	-0.007***	-0.008***	-0.004***	0.039***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.012)
T3: Commitments + Fulf. comuna	-0.014***	-0.006***	-0.008***	-0.010***	-0.011***	-0.006**	0.054***
	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.015)
Competence 2: does what is best for the city							
T1: Commitments	-0.008**	-0.003**	-0.004***	-0.005**	-0.005**	-0.000	0.024***
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.000)	(0.009)
T2: Commitments + Fulf. city	-0.013***	-0.005***	-0.007**	-0.008***	-0.008***	-0.000	0.041***
	(0.005)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.014)
T3: Commitments + Fulf. comuna	-0.012***	-0.004***	-0.007***	-0.007***	-0.007***	-0.000	0.037***
	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.012)
Competence 3: is innovative							
T1: Commitments	-0.003	-0.001	-0.029***	0.032*	-0.029	-0.002	0.031
	(0.007)	(0.008)	(0.009)	(0.017)	(0.019)	(0.025)	(0.026)
T2: Commitments + Fulf. city	-0.011**	-0.003**	-0.005**	-0.008**	-0.006**	-0.001*	0.035**
	(0.005)	(0.001)	(0.002)	(0.004)	(0.003)	(0.001)	(0.016)
T3: Commitments + Fulf. comuna	-0.011	-0.003	-0.005	-0.007	-0.006	-0.001	0.034
	(0.007)	(0.002)	(0.004)	(0.005)	(0.004)	(0.001)	(0.022)
Competence 4: thinks in the long term							
T1: Commitments	-0.004	-0.012	-0.036***	-0.004	0.002	0.000	0.053***
	(0.011)	(0.012)	(0.012)	(0.012)	(0.018)	(0.014)	(0.020)
T2: Commitments + Fulf. city	-0.021***	-0.004***	-0.010***	-0.013***	-0.010***	-0.007***	0.065***
	(0.004)	(0.001)	(0.003)	(0.004)	(0.003)	(0.002)	(0.014)
T3: Commitments + Fulf. comuna	-0.007	0.008	-0.049***	-0.014	-0.035**	0.030	0.067***
	(0.011)	(0.013)	(0.016)	(0.015)	(0.016)	(0.020)	(0.022)
Competence 5: plans and informs its plans							
T1: Commitments	0.007	-0.010	-0.045***	0.000	0.004	0.018	0.026
	(0.007)	(0.007)	(0.012)	(0.015)	(0.017)	(0.022)	(0.022)
T2: Commitments + Fulf. city	-0.021***	-0.009***	-0.012***	-0.013***	-0.011***	0.006**	0.061***
	(0.006)	(0.002)	(0.004)	(0.004)	(0.003)	(0.002)	(0.016)
T3: Commitments + Fulf. comuna	-0.014***	-0.006**	-0.008**	-0.008**	-0.007***	0.004**	0.040**
	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.016)

Panel B: Benevolence

	1	2	3	4	5	6	7
The CABA Government...	**Strongly disagree**	**Disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Agree**	**Strongly agree**
Benevolence 1: acts in the interests of neighbors							
T1: Promises	0.010	-0.025**	0.008	0.012	-0.034	0.023	0.006
	(0.012)	(0.011)	(0.016)	(0.018)	(0.021)	(0.030)	(0.020)
T2: Commitments + Fulf. city	-0.010**	-0.003*	-0.004**	-0.005*	-0.004**	0.004*	0.022**
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.010)
T3: Commitments + Fulf. comuna	0.020	-0.015	-0.024**	-0.027*	0.012	0.016	0.018
	(0.012)	(0.010)	(0.011)	(0.015)	(0.016)	(0.024)	(0.015)
Benevolence 2: does everything in its power to help those in need							
T1: Promises	-0.022***	-0.005**	-0.007**	-0.009**	0.001	0.010***	0.033**
	(0.008)	(0.002)	(0.003)	(0.004)	(0.001)	(0.003)	(0.013)
T2: Commitments + Fulf. city	-0.029***	-0.007**	-0.010***	-0.011***	0.001	0.013***	0.043***
	(0.010)	(0.003)	(0.003)	(0.004)	(0.001)	(0.004)	(0.015)
T3: Commitments + Fulf. comuna	-0.026***	-0.006***	-0.009***	-0.010***	0.001	0.011***	0.038***
	(0.006)	(0.002)	(0.002)	(0.003)	(0.001)	(0.003)	(0.010)
Benevolence 3: pursues policies and projects that my family cares about							
T1: Promises	-0.019***	-0.008**	-0.005***	-0.009***	-0.002	0.008***	0.035***
	(0.007)	(0.003)	(0.002)	(0.003)	(0.001)	(0.003)	(0.013)
T2: Commitments + Fulf. city	-0.014**	-0.006*	-0.004**	-0.007**	-0.002	0.006*	0.026**
	(0.007)	(0.003)	(0.002)	(0.003)	(0.001)	(0.003)	(0.013)
T3: Commitments + Fulf. comuna	-0.028***	-0.011***	-0.007***	-0.013***	-0.003	0.012***	0.051***
	(0.007)	(0.003)	(0.002)	(0.004)	(0.002)	(0.003)	(0.013)

Panel C: Honesty

	1	2	3	4	5	6	7
The CABA Government...	**Strongly disagree**	**Disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Agree**	**Strongly agree**
Honesty 1: is sincere							
T1: Commitments	-0.017***	-0.006***	-0.007***	-0.006***	-0.003**	0.003**	0.036***
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.013)
T2: Commitments + Fulf. city	-0.011*	-0.004	-0.004	-0.004	-0.002*	0.002	0.023
	(0.007)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.014)
T3: Commitments + Fulf. comuna	-0.011*	-0.004*	-0.004*	-0.004*	-0.002*	0.002**	0.024*
	(0.006)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.013)
Honesty 2: is transparent							
T1: Commitments	-0.012**	-0.004*	-0.005*	-0.004*	-0.003*	0.002*	0.027*
	(0.006)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)	(0.014)
T2: Commitments + Fulf. city	-0.013**	-0.004**	-0.005**	-0.005**	-0.003**	0.002*	0.028**
	(0.005)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.012)
T3: Commitments + Fulf. comuna	0.005	-0.006	-0.011	-0.030***	0.024	0.017	0.001
	(0.012)	(0.009)	(0.008)	(0.009)	(0.016)	(0.011)	(0.014)
Honesty 3: fulfills its promises							
T1: Commitments	-0.020***	-0.003***	-0.009***	-0.010***	-0.007***	0.005***	0.044***
	(0.005)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.009)
T2: Commitments + Fulf. city	-0.029***	-0.004***	-0.013***	-0.014***	-0.010***	0.007***	0.063***
	(0.005)	(0.001)	(0.003)	(0.004)	(0.002)	(0.003)	(0.011)
T3: Commitments + Fulf. comuna	-0.020***	-0.003**	-0.009***	-0.009***	-0.007***	0.005**	0.043***
	(0.005)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.010)

Panel D: Trustworthiness (Direct measure)

	1	2	3	4	5	6	7
The CABA Government...	**Strongly disagree**	**Disagree**	**Somewhat disagree**	**Neutral**	**Somewhat agree**	**Agree**	**Strongly agree**
Trust 1: is trustworthy							
T1: Commitments	-0.041***	0.013*	0.003	0.013	-0.031***	-0.010	0.053***
	(0.011)	(0.008)	(0.014)	(0.015)	(0.011)	(0.017)	(0.016)
T2: Commitments + Fulf. city	-0.018***	-0.004**	-0.006***	-0.006***	-0.004**	0.001	0.037***
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.012)
T3: Commitments + Fulf. comuna	-0.013**	-0.003*	-0.005***	-0.005***	-0.003*	0.000	0.027**
	(0.006)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.011)
Observations	2,184	2,184	2,184	2,184	2,184	2,184	2,184

Notes: Clustered standard errors at the Comuna level are shown in parenthesis. Control variables include: age, gender, socio-economic level, labor status, pre-treatment beliefs on government quality and pre-intervention preferences for public education and infrastructure. *p<0.10, **p<0.05, ***p<0.01.

Table 5: Treatment Effect on Trust in Government Conditional on Previous Knowledge of Transparency Initiative and Perceived Government Quality

Category Degree of Familiarity and Perceived Quality of Government (level)	Trust in Government Global Index	Trustworthiness Direct Measure
Familiar, High Quality: T	-0.083	-0.030
	(0.052)	(0.054)
Familiar, Medium Quality: $T + T\ x\ M$	0.201	0.277*
	(0.170)	(0.143)
Familiar, Low Quality: $T + T\ x\ L$	-0.226	-0.327
	(0.341)	(0.320)
Somewhat Familiar, High Quality: $T + T\ x\ S$	0.091**	0.104**
	(0.036)	(0.037)
Somewhat Familiar, Medium Quality: $T + T\ x\ S + T\ x\ M + T\ x\ M\ x\ S$	0.252***	0.276***
	(0.078)	(0.075)
Somewhat Familiar, Low Quality: $T + T\ x\ S + T\ x\ L + T\ x\ L\ x\ S$	0.367**	0.468***
	(0.145)	(0.120)
Unfamiliar, High Quality: $T + T\ x\ U$	0.036	0.023
	(0.064)	(0.075)
Unfamiliar, Medium Quality: $T + T\ x\ U + T\ x\ M + T\ x\ M\ x\ U$	0.174**	0.055
	(0.074)	(0.089)
Unfamiliar, Low Quality: $T + T\ x\ U + T\ x\ L + T\ x\ L\ x\ U$	0.079	0.076
	(0.092)	(0.115)

Notes: T stands for treatment with an informational vignette, *F, S,* and U stand for familiar, somewhat familiar and unfamiliar with the transparency initiative, and *H, M,* and L stand for high (8 to 10), medium (4 to 7) and low (1 to 3) perceived quality of government. This table is based on the results found in Table A8 in the Supplementary Material. We computed standard errors for linear combinations of the estimated parameters; they are shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 6: Calibration of Treatment Effects on Trust

Category	All			Unfamiliar			Somewhat Familiar			Familiar		
	C	T	P $q_{nh,high} = 0.14$	C	T	P $q^{U}_{nh,high} = 0.11$	C	T	P $q^{S}_{nh,high} = 0.20$	C	T	P $q^{F}_{nh,hihg} = 0.10$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	0.149	0.129	0.131	0.207	0.207	0.188	0.108	0.075	0.089	0.131	0.114	0.120
2	0,068	0.058	0.072	0.087	0.093	0.093	0.069	0.044	0.067	0.034	0.032	0.041
3	0.075	0.070	0.070	0.100	0.092	0.094	0.065	0.056	0.061	0.055	0.062	0.050
4	0.131	0.126	0.116	0.170	0.161	0.156	0.137	0.125	0.108	0.055	0.077	0.055
5	0.143	0.145	0.146	0.137	0.159	0.148	0.159	0.162	0.159	0.124	0.094	0.115
6	0.179	0.189	0.187	0.183	0.126	0.187	0.209	0.245	0.219	0.117	0.176	0.125
7	0.253	0.284	0.278	0.116	0.162	0.135	0.253	0.293	0.296	0.483	0.444	0.494
Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Proportion of squared deviations explained			0.791			0.170			0.647			0.071

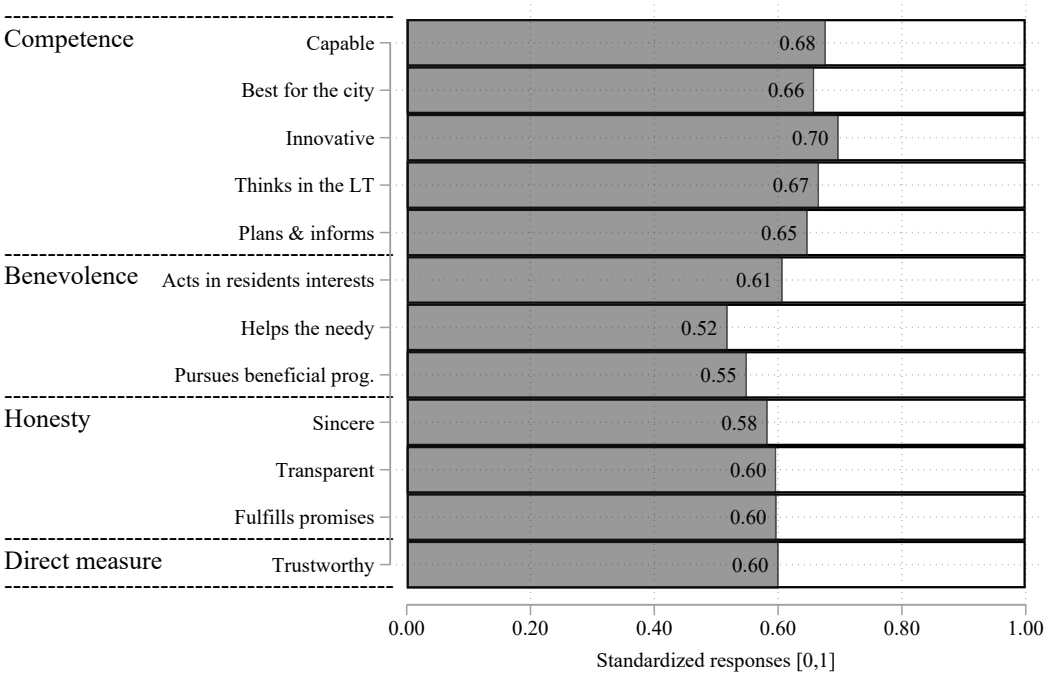
Notes: The direct measure of Trustworthiness is used. There are seven categories of trust that go from 1 (low) to 7 (high). C stands for Control, T for Treatment, and P for Prediction. The last line presents the proportion of the sum of the squared differences between the treatment group and the prediction that can be explained by using the parameter $q_{nh,high}$ that minimizes the sum of squared errors.

Table 7: Potential Effects of Transparency Initiative on Trust

Category	Unfamiliar		Somewhat Familiar		Familiar	
	C	T P $q^U_{nh,high} = 0.11$	C P $q^S_{nh,high} = 0.43$	T P $q^S_{nh,high} = 0.61$	C P $q^F_{nh,high} = 0.77$	T P $q^F_{nh,high} = 0.77$
	(1)	(2)	(3)	(4)	(5)	(6)
1	0.207	0.188	0.126	0.089	0.054	0.054
2	0.087	0.093	0.108	0.107	0.069	0.069
3	0.100	0.094	0.071	0.062	0.088	0.088
4	0.170	0.156	0.094	0.076	0.085	0.085
5	0.137	0.148	0.182	0.138	0.100	0.100
6	0.183	0.187	0.197	0.216	0.178	0.178
7	0.116	0.135	0.223	0.312	0.428	0.428
Total	1.000	1.000	1.000	1.000	1.000	1.000
Proportion of squared deviations explained		0.170	0.836	0.858	0.899	0.954

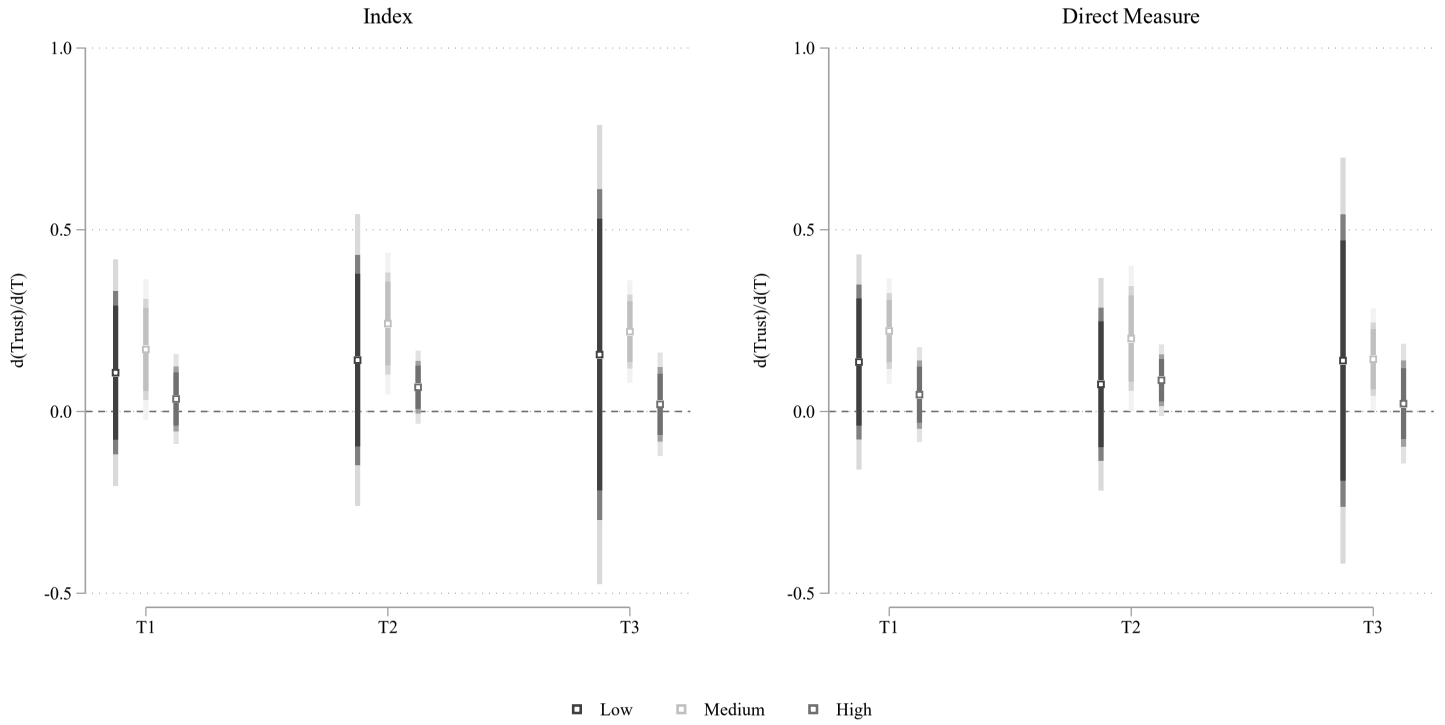
Notes: The direct measure of Trustworthiness is used. There are seven categories of trust that go from 1 (low) to 7 (high). C stands for Control, T for Treatment, and P for Prediction. The last line presents the proportion of the sum of the squared differences between the treatment group and the prediction that can be explained by using the parameter $q_{nh,high}$ that minimizes the sum of squared errors.

Figure 1: Components of the Index: PCA



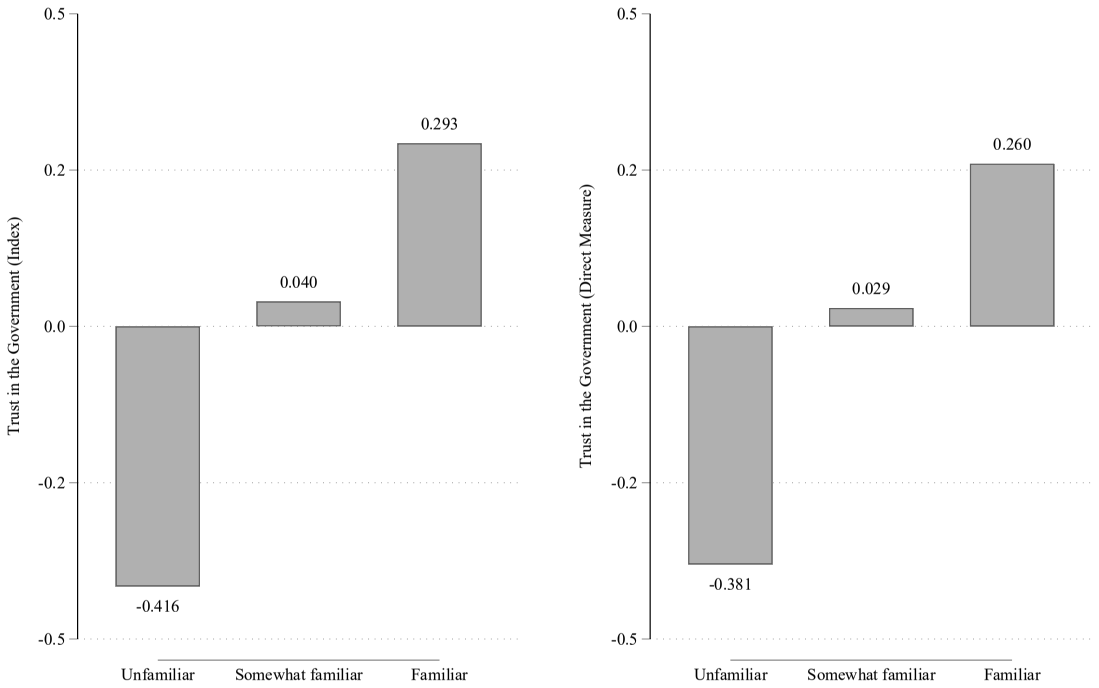
Notes: The figure presents categorical variables that assess citizens' perception of the Government, standardized between zero and one. They account for each of the characteristics asked in the question: *Using a scale from 1 to 7, where one is "Completely disagree," and seven is "Completely agree," please show your level of agreement with the following statements about the Government of the city of Buenos Aires.* The interpretation of each bar goes as follows, e.g., *68%* of the surveyed individuals consider that the CABA Government *is capable*.

Figure 2: Treatment Effect on Trust in the Government by Perceived Quality of the Government



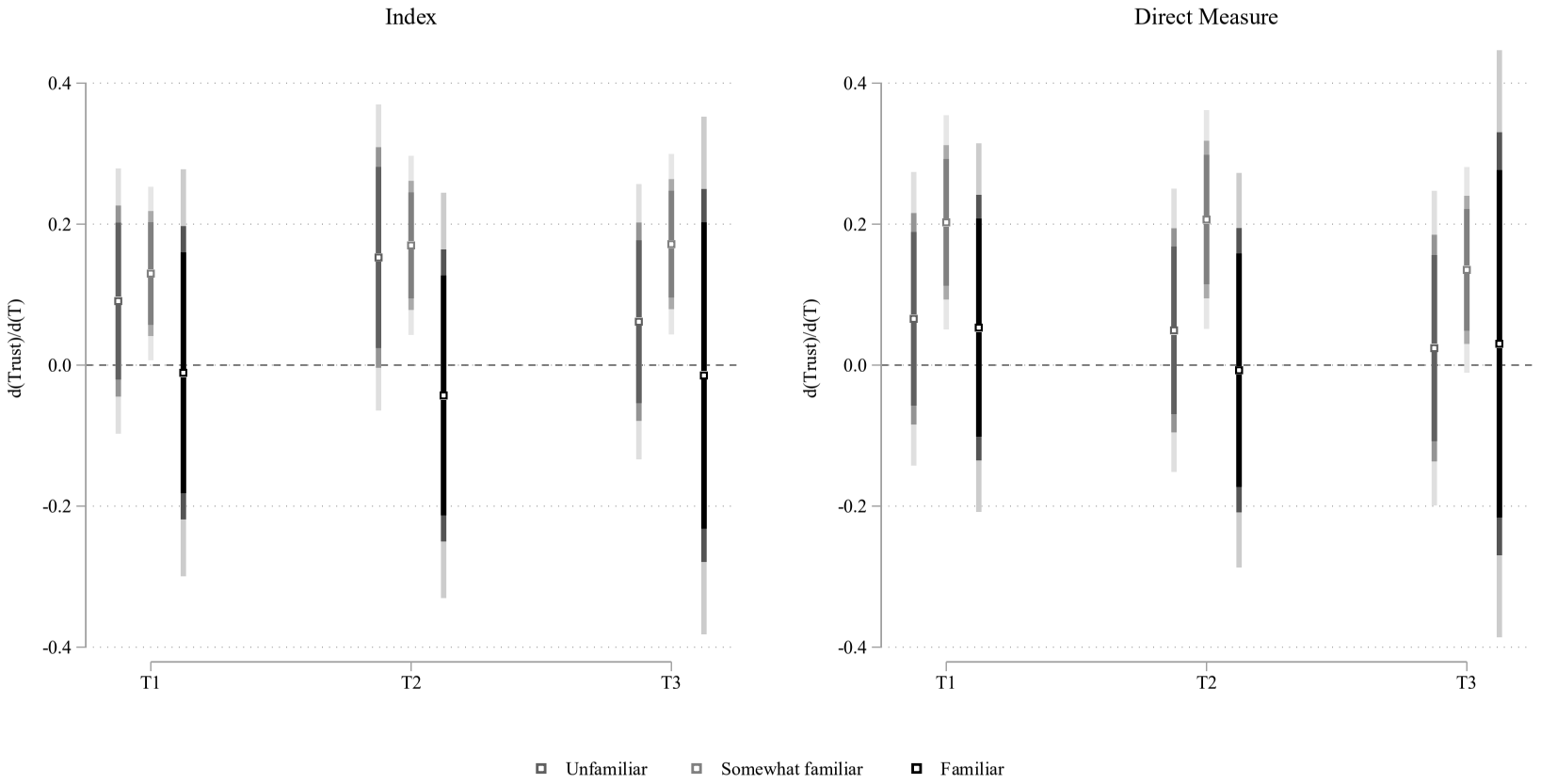
Notes: Perceived quality of the government is a categorical variable that takes values 1 to 10, with one meaning very low quality and 10 very high quality. We reclassified the participants assessments in three categories, low, med-high and high, given the sample composition. People who scored the government quality below 4 are part of the "Low" category, people who scored 4 to 7 in the "Mid" and people who scored 8 or more in the "High" category. The color intensity of confidence intervals represents the confidence level, from darker to lighter, 90%, 95%, and 99%.

Figure 3: Trust in the Government in the Control Group by Previous Knowledge of the Policy



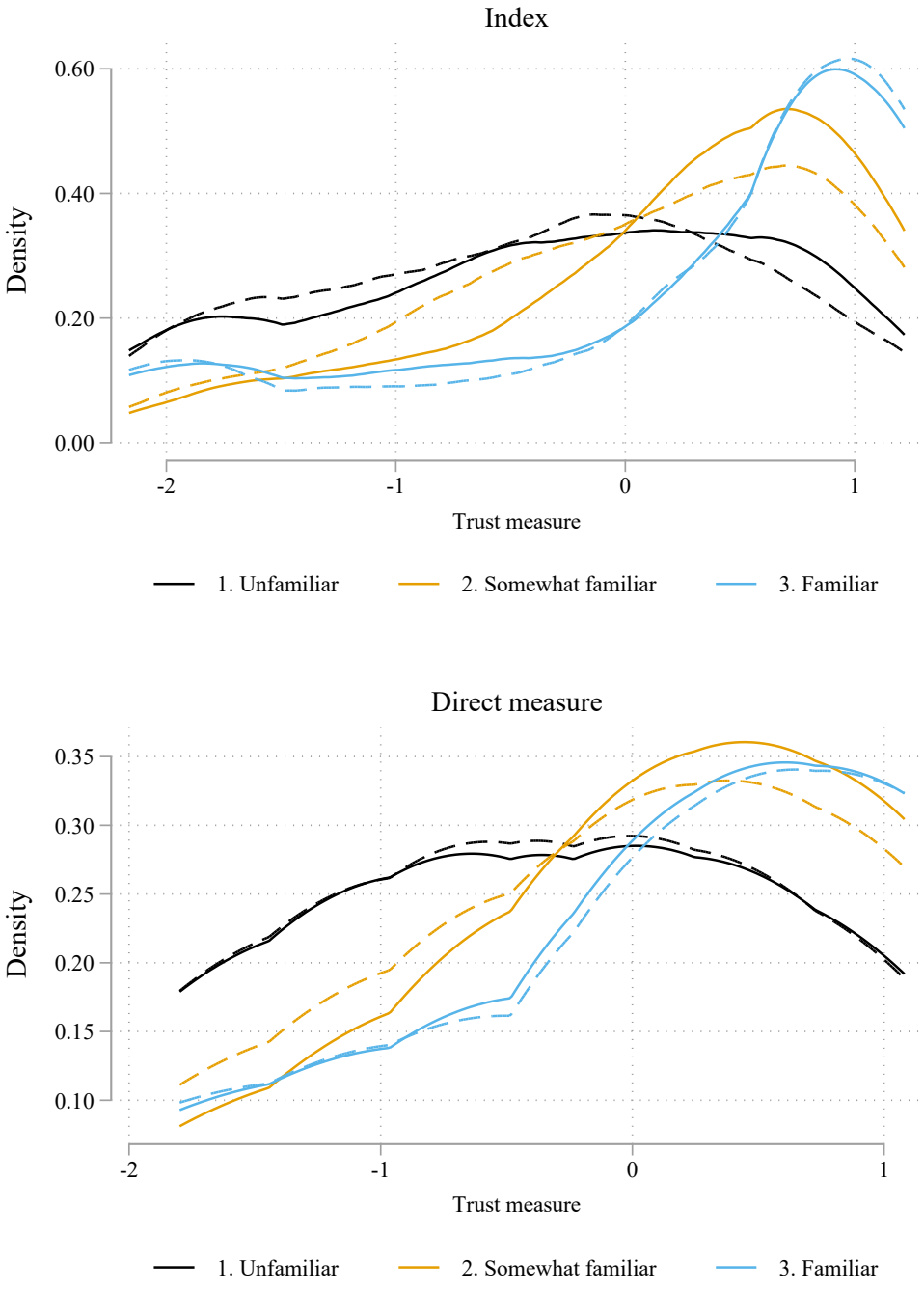
Notes: Dependent variables depicted in this figure are constructed using a PCA method, and standardized with mean zero and standard deviation one. The bars depict the level of trust in the government individuals from the control group have grouped by level of initial knowledge of the commitments policy. They could be classified as *unfamiliar* if they did not know the policy at all, *somewhat* familiar if they had heard about it, or *familiar* if they were aware of it beforehand.

Figure 4: Treatment Effect on Trust in the Government by Level of Knowledge of "Compromisos"



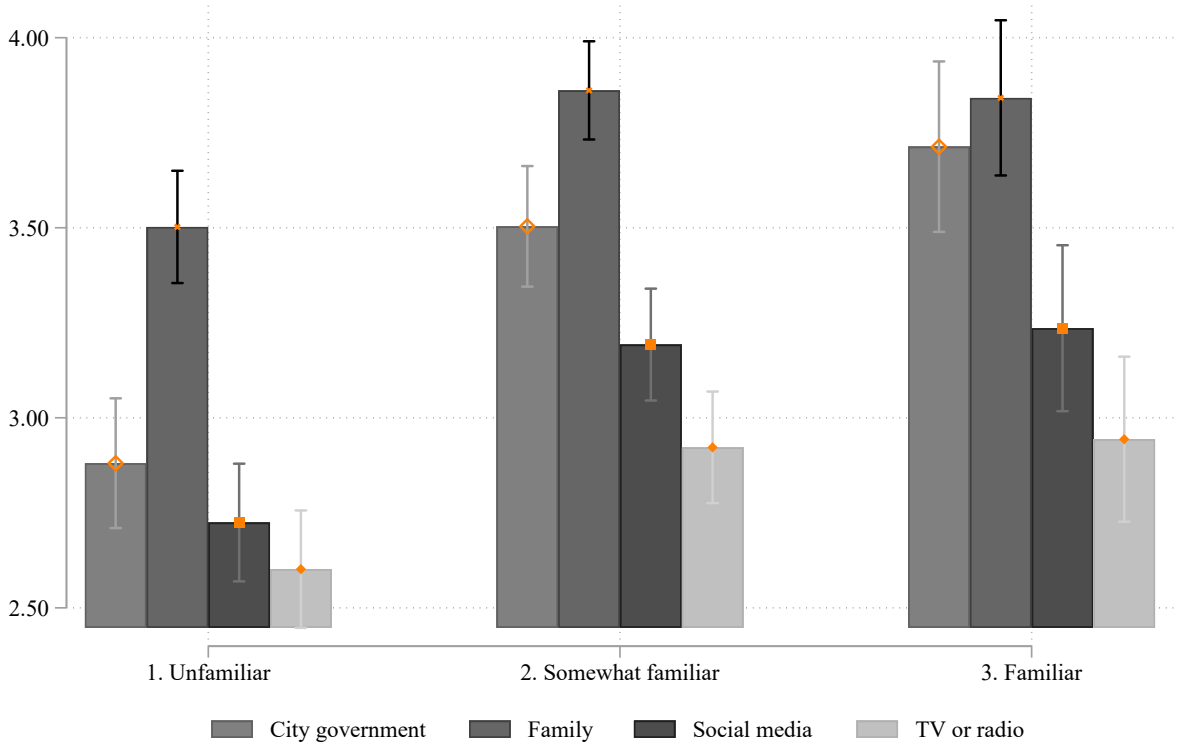
Notes: Dependent variables depicted in this figure are constructed using a PCA method, and a standardization with mean zero and standard deviation one. The estimate points correspond to the effect of each treatment in the level of knowledge indicated in the legend. For example, people who had heard of the commitments but did not know them for sure (somewhat familiar) before the survey increase their trust in the government by a greater extent than people who did not know them at all (unfamiliar) or those who knew the policy (familiar), regardless of the trust measure used. The color intensity of confidence intervals represents the confidence level, from darker to lighter, 90%, 95%, and 99%.

Figure 5: Distribution of Trust by Level of Knowledge of "Compromisos"



Notes: Dashed lines (Dd) correspond to the distribution of the Trust measure for the control while solid lines (Ds) for the treated group. We present two-sample Kolmogorov-Smirnov tests for equality of the distributions in a pairwise fashion for the **Index measure**, Direct Trust available upon request. Comparisons of Ds,k vs. Dd,k—solid versus dashed lines for each knowledge level k—are uni-directional and one-tailed tests, where $k = 1$. They indicate whether the trust index is smaller for the Dd (control group) compared to Ds (treatment). Ds,k=1 vs. Dd,k=1: 0.079 (0.236); Ds,k=2 vs. Dd,k=2: 0.120 (0.006); Ds,k=3 vs. Dd,k=3: 0.073 (0.621). Finally, we include the test for people who were treated and already knew about the commitments and those who had only heard about them to test for whether the impact of the information can level trust between those who already knew and those who acquire new information. Ds,k=3 vs Ds,k=2: 0.169 (0.000)

Figure 6: Average Trust in Information Provided by Different Agents, by Level of Previous Knowledge (control group)



Notes: All variables are categorical and respond to the question "To what extent do you trust the information about government performance and management provided by [*the Government / your family / social media / traditional media (TV or radio)*]?" They take values between one and five, 1 being the lowest category ("Not at all"), and 5 the highest ("Very much").