LEMMA Working Paper

n° 2025-09

Externalities and Distrust

Gianmarco Daniele

University of Milan and Bocconi University

Andrea F.M. Martinangeli

Université Paris-Panthéon-Assas, LEMMA

Francesco Passarelli

University of Turin, Collegio Carlo Alberto, CESifo and Baffi Carefin at Bocconi

Willem Sas

Hasselt University, University of Stirling, CESifo and KU Leuven

Lisa Windsteiger

University of Salzburg, and ifo Institute

LEMMA EA 4442 - 4 rue Blaise Desgoffe 75006 Paris

Externalities and Distrust*

Gianmarco Daniele[†] Andrea F.M. Martinangeli[‡]
Francesco Passarelli[§] Willem Sas[¶] Lisa Windsteiger[‖]

September 10, 2025

Abstract

We present a theory explaining how divergent opinions about the regulation of an externality problem can lead to distrust of both political institutions and other people. Those who are more concerned about the externality distrust the government because they perceive the rule as too lenient. Those who are less concerned blame the government because they have to comply with a rule they consider useless.

In addition, the rule shapes individuals' expectations of civic behavior in society. Therefore, regulation can lead to disappointment and interpersonal distrust when

*We are grateful to Marco Battaglini, Tito Boeri, Davide Cantoni, David Card, Pierluigi Conzo, Stefano DellaVigna, Martin Dufwenberg, Roberto Hernan Gonzalez, Doug Irwin, Alexander Koch, Nicola Maaser, Giovanni Maggi, Armando Miano, Massimo Morelli, Salvatore Nunnari, Pietro Ortoleva, Amedeo Piolatto, Diego Ramos-Toro, Pilar Sorribas-Navarro, Eli Spiegelman, Robert Staiger, Stefanie Stantcheva, Guido Tabellini, Mathias Trabandt and Francesco Trebbi for their helpful comments. This work was presented at several venues (Glasgow, St Andrews, Munich-LMU, Lisbon, Fürth, Bari, Rome-Sapienza, Turin, Paris-Sorbonne, Aarhus, Rotterdam): we thank organisers and attendants. Financial support from the Max Planck Society, the ISITE-UBFC International Coach Fellowship (ANR-15-IDEX-0003) and Italian MIUR (PRIN 2020PLZR2P and GRINS PE00000018) and the grant PID2022-137707NB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER, UE is gratefully acknowledged. An earlier version of this paper titled "Externalities and Distrust" was awarded the 'Bank of Portugal best paper prize' at the 2022 Lisbon Meeting on Economics and Public Science.

[†]University of Milan and Bocconi University
[‡]Université Paris 2 Panthéon-Assas, France
[§]University of Turin, Collegio Carlo Alberto, CESifo, and Baffi Carefin at Bocconi.
[¶]Hasselt University, University of Stirling, CESifo, and KU Leuven
[‖]University of Salzburg, and ifo Institute.

others do not follow the rules. Those who are more concerned also value civic behavior more, so they lose even more trust in others. Our survey experiment in four European countries supports these patterns in the context of the Covid-19 pandemic.

JEL Classification: D70, D72, H3, O52.

Keywords: Social Trust, Institutional Trust, Heterogeneity, Externalities, Regulation, Survey Experiment, Covid-19, Climate Change.

1 Introduction

Trust in political institutions is a fundamental ingredient for the stability of democracies and the effectiveness of government action. However, political trust has been declining in nearly all Western countries over the last decades.[1]

At the same time trust in others is also rather low, as people are seen as less reliable than they used to be.[2]

Sadly, whenever societies face a common challenge and the need for trust is highest (Ostrom, 1998), citizens seem to become critical of the government's ability to implement the right policies and of people's ability to comply with those policies.

Research in sociology, political science, and economics finds that high levels of well-being and social progress are associated with high levels of trust in others and in institutions (e.g. Putnam (1993); Knack and Keefer (1995); Guiso et al. (2006); Tabellini (2008), among many others). But why do advanced democracies struggle to maintain high levels of institutional and interpersonal trust? In this paper we present a theory showing how these two forms of distrust are determined and how they are linked. We then conduct a survey experiment providing support for our theory.[3]

Our starting point is the idea that regulation can disappoint citizens, and hence stir up distrust. In a complex and interdependent world, the government increasingly relies on regulation to encourage cooperation between economic agents in order to mitigate

[1] According to Pew Research Center, the share of Americans saying that *they trust the government to do what is right just about always/most of the time* has dropped from 75% in 1958 to 20% in 2022. In the European Union, levels of institutional trust vary significantly across countries. Eurobarometer data show that on average 4 out of 10 people distrust their national government, with an increasing trend in the last two decades and an acceleration in the post-Covid period. Similar patterns are observed in OECD countries, where trust surveys document an equal split between those who trust and those who distrust their political institutions.
https://www.pewresearch.org/politics/2022/06/06/public-trust-in-government-1958-2022/
https://www.oecd.org/governance/trust-in-government/

[2] Putnam et al. (2000) documents a sharp decrease in social trust in the US between 1960 and 2000. Using integrated social surveys, Our World in Data has recently shown that the percentage of people who agree with the statement *Most people can be trusted* is only 37% in the USA. This percentage is slightly higher than 40% in Spain, Germany, and UK. It is 26-27% in Italy and France, while it drops to much lower levels in most Eastern European countries. In July 2022, the Pew Research Center reported that 71% of Americans believed that interpersonal trust has worsened over the past 20 years. Most of them thought that this was due to political polarization and lack of political trust.
https://www.ourworldindata.org/grapher/self-reported-trust-attitudes/

[3] Throughout this paper, we use the term "social trust" interchangeably with "generalized" or "interpersonal" trust, and the term "political trust" interchangeably with "institutional" trust. In more specific terms we focus on trust in political institutions as well as generalised trust (in unknown others), as defined in subsection 2.4 where we discuss how our conceptualisation aligns with existing definitions in the literature.

negative externalities.[4] At the same time, individuals have their own expectations about the rules that the government should adopt. If they think a rule is wrong, they lose trust in political institutions. Moreover, once a rule has been adopted it becomes a standard of civic behavior, and hence a reference point to judge others. When individuals see that the rule is not respected, they will also lose trust in others as a result.

What matters, then, is the degree to which people disagree. We assume that individuals hold different beliefs about the severity of an externality problem, and hence will have different opinions about what the government should do. For the same reason, they also have different incentives to comply with government policies. As a result, greater heterogeneity in beliefs increases the risk of trust erosion in both institutions and others. This *heterogeneity* in beliefs may stem from different ideological and cultural traits, varying material interests, different information or ability to process it, or even different levels of exposure to social media. Although all of these mechanisms are interesting *per se*, our focus in this paper lies with the next stage. Once beliefs are formed, we are interested in what happens when citizens compare their subjectively optimal policies with actual policies, and the resulting expectations of civic behavior with actual behavior of others.

In order to focus on how distrust is determined, we simply assume the government makes its decisions based on its own view of the severity of the externality. We remain agnostic about where that view comes from and abstract from political constraints or distortions. Once the rule is in place, those who are more concerned about the externality would have preferred more restrictive rules and are disappointed because the government does too little. They are, however, still more willing to abide by the rule, even when deemed too lenient. They also believe that those who do not follow the rule harm them considerably. On the other hand, individuals who care less about the externality will distrust the government for setting a policy they believe is too intrusive, whilst they consider misbehavior of others not too harmful. As they often don't fully comply themselves, they will only distrust those who comply less than them.

Our theory shows that greater heterogeneity in beliefs increases the possibility of trust erosion, both in institutions and in others. Moreover, heterogeneity of opinions yields *asymmetric* expressions of institutional and interpersonal distrust. Concerned individuals develop stronger feelings of interpersonal distrust because they think the irresponsible

[4]Taxation would be an alternative to regulation. Here we focus on regulation because, besides distortionary effects when it reaches high levels, taxing behavior is not always technically feasible.

behavior of non-complying individuals causes substantial social harm. On the other hand, when mitigating behavior imposes substantial costs, less concerned individuals develop feelings of distrust mainly towards institutions, because they think the government is uselessly and expensively restricting their freedom. This "divergence" in trust, a dimension so far neglected, seems quite realistic and might help explain - among other effects - trust polarization in societies, with "disobedient" groups mostly distrusting political institutions and "civic" groups mostly distrusting disobedient individuals.

Our empirical analysis suggests that the divergence in trust described by the model arises when policy issues are highly salient and mired in coordination failures. We conducted a large online survey experiment in four European countries in the aftermath of the first lockdown periods in 2020. We find negative correlations between self-reported health concern about Covid-19 and the view that regulations were too strict, indicating that the non-concerned indeed thought the rules were too intrusive or ineffective. We also find positive correlations between health concern and self-reported compliance with emergency regulations, the expected cooperative behavior of others, and appreciation of government performance during the pandemic. In line with our theory, this finding suggests that concerned individuals were indeed more demanding of others yet more lenient towards the government, whilst themselves following the rules. Moreover, we find that, when primed with several salient aspects of the crisis, individuals who said they were more worried about the virus lost a lot more trust in others than in institutions, compared to those saying they were less worried. It is hence plausible that, once led to thinking intensely about Covid-19 by our priming treatments, individuals focused on the sizable economic costs associated to lockdown measures and on the strictness of the rules. As a result, the more worried ones were more disappointed by the non-compliant behavior of others, while the less worried were more disappointed by the government policy. These results seem realistic and in line with the Vax vs No-Vax diatribe of those days.

Our theoretical framework yields several insights. First, the more diversified opinions are, the more social and political trust is lost. The increasing diversity of opinions, eventually fueled by social media (Gentzkow, 2016), or by "polarization of reality" (Alesina et al., 2020; Stantcheva, 2021), erodes trust within society.

Second, the greater the interests at stake, and hence the more pronounced the externality, the greater the loss of interpersonal and political trust in society. This situation

would lead to increased social tension and polarization, particularly during times of crisis, a result that is consistent with pro-cyclical patterns of distrust uncovered by the empirical literature (Stevenson and Wolfers, 2011).

Third, when new challenges need government intervention, disagreement on what actions the government should take and misaligned expectations on how people should behave yield distrust. Therefore, while our static framework cannot determine whether these effects are persistent or not, and what rebuilds trust, it may help explain why trust has been declining in many countries over the last decades.

In our framework, civic/uncivic behavior and interpersonal trust are *endogenously* determined by regulation. In this respect our paper is close to Aghion et al. (2010) who posit that regulation itself breeds distrust. They also show how how distrust brings about the demand for regulation (as in Algan and Cahuc (2009) and Pinotti (2012)). Martinangeli et al. (2024) show that institutions leaving room for undetectable corrupt behavior causes sharp drops in social trust. In these papers, corruption and bribery are the mechanisms at play. Here, we focus on differences in policy opinions among people, which enables us to directly study asymmetric forms of distrust and to propose a novel link between trust in institutions and interpersonal trust.

Our model shares common features with the work of Gitmez et al. (2020). Similarly to them we consider a framework where individuals hold heterogeneous preferences on how to solve an externality problem. While they focus on strategic information acquisition and voting, we focus on trust erosion.

Our work is related and in some respects complementary to the work of Durante et al. (2021) and Algan et al. (2021). They study compliance and trust during the Covid-19 pandemic. While we show how low compliance and low policy support lead to distrust, they show how distrust can feed back into compliance and policy support, triggering a vicious circle.

The key role of preference heterogeneity in our theory links this paper to the literature studying the relationships between diversity in norms, values or preferences and political-economy outcomes, such as civic cohesion or the provision of public goods (e.g. Alesina and Ferrara (2005); Alesina et al. (2016)). Alesina et al. (2017) find large and increasing differences in values, trust, and other cultural traits between and within European countries. They conjecture that sharing common policies in the EU may have bred

conflicts and antagonized public opinions. Our theory is consistent with their conjecture.

While most of the political economy literature studies how heterogeneity of preferences affects collective decisions, in this paper we study what happens when the collective decision, once made, does not reflect the preferences of individuals. Although this can be a very important problem in democracies, it is relatively understudied. In exploring it, we are close to the work of Passarelli and Tabellini (2017) who examine the impact of political disappointment on protests. We expand on their idea that individuals compare the actual policy with a counterfactual policy that serves as a fair reference point. Here, we use two counterfactual outcomes to size up on individuals' lack of trust in the government and on their lack of trust in others. The crucial link between these two counterfactual reference points, allows us to show how distrust in the government and distrust in others may be related.

Our theory helps rationalize the trust erosion that we observe in our experimental investigation of the Covid-19 pandemic. However, the theory is general enough to be applied, sometimes with some caveats, to a wide variety of externality problems, ranging from traffic rules, to safety standards, to construction regulations. After the pandemic, the next major "externality challenge" that nations will face is global warming. Opinions on climate change are highly heterogeneous and may become even more diverse in the future (Canen et al., 2020). Our model predicts a generalized loss of trust in this scenario.

Finally, it is important to highlight that our approach is not normative. Nevertheless, our model can shed light on the consequences of policy interventions on distrust and help design policies that can stem the erosion of trust. We will return on this point in the conclusions.

2 A General Model of Externalities and Trust

Consider a unit-sized society with a continuum of individuals i uniformly distributed in the interval $[0, 1]$. Sometimes it will be useful to refer to "individual i" or "type i", and later it will become clear which individual trait is captured by i.

Let $b_i \in [0, \infty)$ denote individual behavior. Increasing b_i yields private gains but also social benefits. The Covid-19 crisis provides an intuitive example of what we have in mind: increasing social distance, wearing a mask more frequently or getting a vaccine

reduces the risk that individual i gets infected (the private benefit), but also mitigates the risk that other individuals get infected or have to pay the cost of i's hospitalization (the social benefit). Another example is climate change. Environmentally friendly behavior lowers the energy bill, but also reduces emissions. Besides these two examples, our model captures many realistic situations in which externalities yield social coordination problems.

We are thus thinking of a "virtuous" dimension in people's behavior that not only provides private benefits but also mitigates a negative externality (or prompts a positive one).[5] Any individual i will benefit from her own mitigating behavior, b_i, and from that of others. To capture this idea, let her benefit function be

$$\varepsilon_i \left(b_i + B_{-i} \right), \tag{1}$$

where B_{-i} is the average behavior of others

$$B_{-i} \equiv \int_0^1 b_j dj, \qquad \forall j \neq i, \tag{2}$$

and $\varepsilon_i(.)$ a twice differentiable function with $\varepsilon_i'(.) > 0$ and $\varepsilon_i''(.) \leq 0$. Here, for simplicity, we are assuming that the behavior of others enters i's benefit function symmetrically w.r.t. b_i, and it is captured by their unweighted average, B_{-i}.[6]

Let us now introduce heterogeneity across individuals. We assume they are different in the perception of how their behavior (and the behavior of others) affects their welfare. This is realistic in many situations. For instance, we discussed in the introduction that concern about Covid-19 is highly heterogeneous across people. The strongly concerned will value mitigation highly as they think the virus is very infectious, whilst others who

[5]Mitigating a negative externality or prompting a positive one are formally equivalent in this model. Hereafter, by thinking of the negative externality, we will often refer to increasing behavior as "mitigating behavior".

[6]More realistically i might be affected more by her own behavior, or by that of some individuals (e.g. her neighbors or relatives) than others, and with different weights. Qualitatively, however, the message of this model would not change substantially if we were to account for different degrees of 'proximity' to others. This model can also capture situations in which individuals' behavior yields pure externalities with no private gains. In this case the benefit function would be $\varepsilon_i \left(B_{-i} \right)$. Note that concavity of the benefit function captures the idea that one's incentive to mitigate behavior is lower at the margin when other people are already mitigating theirs, as in the case of social-distancing in a pandemic. One might also assume convexity. For instance an individual's incentive to use, say, the train instead of her car increases at the margin when more people are using the train and more options are available in the train timetable. It will become clear that concavity (convexity) implies strategic substitutability (complementarity).

believe that Covid-19 is a simple cold, or even deny the virus exists, value mitigation less.

We capture concern through the first derivative of (1). The idea is that a more concerned individual values the mitigating effect of behavior at the margin more. We say that i is more concerned than j if $\varepsilon_i'(y) > \varepsilon_j'(y)$, for any value of y.[7] By our simplifying assumption that $i \sim U[0,1]$ it follows that the degree of concern is uniformly distributed across the population. Nothing crucial hinges on this assumption.[8] In order to introduce a spatial intuition, and hence to benefit from integral calculus, we assume a positive monotonic relationship between types and degree of concern, $\frac{\partial \varepsilon_i'(y)}{\partial i} > 0$ for any y.

Mitigating behavior comes at a private cost expressed by an increasing function $c(b_i)$, e.g. the time cost of having to find and put on a mask, the effort of social distancing, the cost of losing business opportunities because of lockdown, but also the cost of a solar panel, of having to sell an old car and buying a new less-polluting one, etc. We assume the cost function is convex and twice differentiable. Because we study how individuals' preferences and choices are driven by their *perceptions* of the private and social value of mitigating behavior, we assume identical costs across the population. In other words, individuals are heterogeneous only in their degree of concern about the effect of behavior, while they are identical in all other respects.[9] For simplicity, we also assume that $c'(0) = 0$.

Combining benefits and costs, individual utility then becomes

$$U_i(b_i, B_{-i}) = \varepsilon_i(b_i + B_{-i}) - c(b_i).$$

For all i, optimal *decentralized* behavior maximizes $U_i(b_i, B_{-i})$ by solving

$$\varepsilon_i'(b_i + B_{-i}) = c'(b_i), \tag{3}$$

which has an interior solution as $\varepsilon_i''(.) - c''(.) < 0$. Equation (3) defines i's best response function. Note that, as long as $\varepsilon_i''(.)$ is strictly negative, i's optimal behavior negatively depends on the behavior of others, B_{-i}, and there is strategic substitutability. The resulting Nash-equilibrium is achieved when all individuals are on their reaction functions.

[7] We assume that if $\varepsilon_i'(y) > \varepsilon_j'(y)$ holds for some y, it holds for all y, a sort of 'concern consistency'.

[8] As discussed in the introduction, we remain agnostic as to the underlying mechanisms driving concern, since we focus here on the effect of concern heterogeneity across individuals.

[9] Alternatively one could assume that individuals differ in the effort costs of mitigating behavior. Heterogeneity would then come from the cost side. Such a model would capture different situations, but would lead to similar predictions regarding the political disagreement across people.

Utility in a decentralized equilibrium becomes

$$U_i\left(b_i^*, B_{-i}^*\right) = \varepsilon_i\left(b_i^* + B_{-i}^*\right) - c(b_i^*),$$

where b_i^* is i's decentralized equilibrium behavior, and $B_{-i}^* \equiv \int_0^1 b_j^* dj, \quad \forall j \neq i$.[10] No individual internalizes her positive contribution to other individuals' welfare, and the non-internalised externality leads to sub-optimal levels of mitigation. Indeed, if i were to internalize the effect of her behavior on the entire society, she would face a different optimization problem, leading to more mitigating behavior.[11] This leaves scope for policy intervention to fix the externality problem.

By implicitly differentiating (3) it follows that more concerned types will mitigate more, $\frac{\partial b_i^*}{\partial i} > 0$. This result is not due to any form of social or moral concern, but is entirely driven by stronger private concerns. In other words, in a decentralized equilibrium individuals behave more virtuously only when they are more concerned about the effect of their behavior on their *own* welfare. In this model individuals are not altruistic, as we abstract from morally motivated pro-sociality.[12]

2.1 Behavioral Rules and Enforcement

To incentivize people to engage in socially beneficial behavior, we consider a rule fixing a lower bound ρ to the behavior of all individuals.[13] The government has to set the rule (more on this later) and decide how to enforce it. In this context, it seems realistic for the government to rely on fines and monitoring activities. It might aim for full compliance, which would be ensured by sufficiently high fines and strict monitoring. This goal may be difficult to achieve in practice, however, especially if monitoring individuals' behavior is costly and excessively high fines are considered unacceptable to society for some reason.

[10] Of course there is a background problem of equilibrium existence, stability and uniqueness. For tractability, we assume throughout that a decentralized equilibrium exists, is unique and stable. We make a similar assumption for any other equilibrium concepts we will present.

[11] She would maximize $U_i\left(b_i, B_{-i}\right) + \int_0^1 U_j\left(b_j, B_{-j}\right)$ wrt b_i. The FOC would be $\varepsilon_i'(.) + \int_0^1 \varepsilon_j'(.)\frac{\partial B_{-j}}{\partial b_i} dj = c'(.)$. Because $\int_0^1 \varepsilon_j'(.)\frac{\partial B_{-j}}{\partial b_i} dj = \int_0^1 \varepsilon_j'(.) dj > 0$, the internalization equilibrium would imply a higher mitigating behavior compared to the decentralized one.

[12] Absent private gains, as in the case of a pure externality, equilibrium decentralized behavior would be $b_i^* = 0$, for all i. No individual would make any effort to mitigate the externality.

[13] Of course there are other instruments (e.g. Pigouvian taxes, subsidies, tradeable permits, or combinations of them). Here we consider rules, which are widely used in reality perhaps because they are simple and cheap to implement.

Here we assume that the government imposes the highest "socially tolerable" fine, $\overline{k}$, and that the probability to catch a non-compliant individual positively depends on two factors. First, the efficacy of monitoring, which is parametrized by $m \in [0,1]$, our policy variable. Second, the "degree of non-compliance", defined as the difference between the rule and an individual's behavior, if positive.

Formally, let $p(b_i, \rho, m) \in [0,1]$ be the probability to be caught. If $b_i < \rho$, this probability is positive and increasing in ρ and in m, and decreasing in b_i, while it is zero if $b_i \geq \rho$. We also assume for simplicity it is continuous, twice differentiable, and convex in misbehavior, $(\rho - b_i)^+$, with $\frac{\partial p(b_i,\rho,m)}{\partial b_i} = 0\Big|_{b_i \geq \rho}$. The idea is that stricter control by e.g. the police (i.e., higher m) yields a higher probability to catch someone who does not comply with the rule. Moreover, the more an individual misbehaves the more she becomes "detectable". Finally, to capture the idea that stricter monitoring increases individuals' incentives to mitigate their behavior, we further assume that the marginal impact of behavior on this probability is stronger and sufficiently large if m is stricter and it is independent of ρ.[14]

2.1.1 Rules, enforcement, and equilibrium behavior

Now we can study how individual i optimally reacts to the government's rule ρ and monitoring, m. Her expected utility is

$$U_i(b_i, \rho, m) = \begin{cases} \varepsilon_i\left(b_i + B_{-i}\right) - c(b_i) & \text{if } b_i > \rho \\ \varepsilon_i\left(b_i + B_{-i}\right) - c(b_i) - \overline{k} \cdot p(b_i, \rho, m) & \text{if } b_i \leq \rho \end{cases}. \tag{4}$$

where $\overline{k} \cdot p(b_i, \rho, m)$ is the expected fine when she does not respect the rule. In equilibrium i chooses her behavior so as to maximize (4), given the existence of the rule ρ, the government's monitoring activity, m, and given the equilibrium behavior of all other individuals. So, for all i, non-decentralized equilibrium behavior, $b_i^*(\rho, m)$, satisfies the following optimality condition at the equilibrium point:

$$\varepsilon_i'(b_i^*(.) + B_{-i}^*(.)) - c'(b_i^*(.)) = 0 \quad \text{if } b_i^*(.) > \rho \tag{5}$$

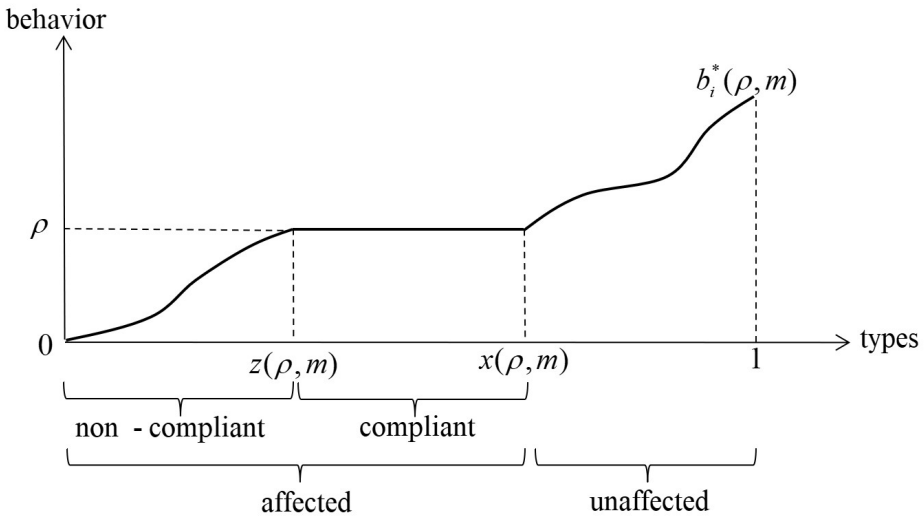
$$\varepsilon_i'(b_i^*(.) + B_{-i}^*(.)) - c'(b_i^*(.)) - \overline{k} \cdot \frac{\partial p(b_i^*(.), \rho, m)}{\partial b_i^*(.)} \geq 0 \quad \text{if } b_i^*(.) \leq \rho \tag{6}$$

[14] Thus, $\frac{\partial^2 p(b_i,\rho,m)}{\partial b_i \partial m} < 0$ and it is sufficiently large in absolute value, and $\frac{\partial^2 p(b_i,\rho,m)}{\partial b_i \partial \rho} = 0$. The proof of Lemma 1 in Appendix A provides further details on this assumption.

where $b_i^*(.) \equiv b_i^*(\rho, m)$.

Equation (5) refers to the group of *unaffected* individuals. They are the most concerned ones, so they voluntarily choose a behavior that is more virtuous than the government's rule, as shown by Figure 1. These individuals do not really care about m because they bear no risk of being fined. Yet, their behavior, $b_i^*(\rho, m)$, indirectly depends on m and ρ because in equilibrium they optimally react to the behavior of the affected ones, which in turn depends on m and ρ.[15] Lemma 1 below states that this group shrinks when the rule is more stringent or monitoring is stricter. More people will comply with the rule and, due to strategic substitutability, less people will choose to behave more virtuously than the rule.

Figure 1: Equilibrium behavior



Expression (6) refers to the *affected* individuals. Somehow interestingly, they split in two sub-groups. First, those who choose to fully obey the rule and bear no risk to pay the fine. They are less concerned than the unaffected, and more concerned than the second sub-group (see below). Figure 1 shows that their fully-compliant behavior is ρ and their types range from type $z(\rho, m)$ to type $x(\rho, m)$. The size of this group enlarges when monitoring is intensified because more people will choose to comply. The second sub-group

[15] Note that those individuals we refer to as "unaffected" are still *indirectly* affected by the rule because they change their behavior in equilibrium in response to the behavior of others in society. The label "unaffected" serves the purpose of distinguishing these individuals from those whose behavior is directly constrained by the rule, i.e. the least concerned, or "affected", types described below.

consists of the least concerned types below $z(\rho, m)$. Monitoring is not sufficient to lead these individuals to fully comply. So they bear the risk of being fined. However, Lemma 1 indicates their behavior is more virtuous than the decentralized behavior without monitoring. The perspective of being fined if caught makes them increase their behavior with the intent of reducing the expected fine. It further indicates that when the rule is stricter less people will comply in equilibrium.[16]

We finally assume that when the effectiveness of monitoring is maximal (i.e., $m = 1$) then everyone chooses to comply in equilibrium:

$$b_i^*(\rho, 1) \geq \rho \qquad \text{for any } i \text{ and any } \rho \tag{7}$$

In other words, (6) is a strict inequality for all i and any ρ when $m = 1$. This assumption is not fundamental, but it is useful because it will make our definitions of individuals' most preferred rules and their interpersonal distrust in Sections 2.2 and 2.4.2, respectively, more transparent. Lemma 1 summarizes these results.[17]

Lemma 1. *In the presence of a government rule, ρ, and monitoring activity, m, society is split in three groups as follows:*

1. *Unaffected individuals, with $b_i^*(\rho, m) > \rho$ and $i \in (x(\rho, m), 1]$. Their equilibrium behavior, $b_i^*(\rho, m)$, is decreasing in government regulation, ρ, and in monitoring policy, m. The lowest individual in this group, $x(\rho, m)$, positively depends on m and ρ.*

2. *Affected individuals deciding to fully comply, with $b_i^*(\rho, m) = \rho$, and $i \in [z(\rho, m), x(\rho, m)]$.*

3. *Affected individuals choosing not to fully comply, with $b_i^*(\rho, m) < \rho$, and $i \in [0, z(\rho, m))$. Their behavior negatively depends on the rule, ρ, and positively on monitoring, m. The highest individual in this group, $z(\rho, m)$, positively depends on ρ.*

4. *By (7) with perfect monitoring all individuals comply, thus $z(\rho, 1) = 0$.*

[16] Note that $z(\rho, m)$ is the highest i such that (6) is an equality while $x(\rho, m)$ is the lowest i such that (5) is an equality.

[17] Proofs of Lemma 1 and all following propositions are provided in Appendix A.

2.2 Individuals' Policy Entitlements

Individuals have different perceptions about the benefits of their behavior and the behavior of others, so they will have different views about what the right rule should be. Let ρ_i^* denote individual i's most preferred rule. Importantly, we assume that any individual chooses ρ_i^* under the expectation that all people will *fully abide* by that rule. This assumption reflects the idea that once any rule has been set by the government, it shapes the notion of civic behavior in the society. All individuals *should* fully abide by that rule, with no need to detect or sanction non-compliant behavior. In a sense, ρ_i^* represents i's moral view on society, which reflects i's "double moral expectation". On the one hand, the government *should* be as concerned as she is and set a rule equal to ρ_i^*. On the other, all individuals *should* care about the consequences of their behavior on others and follow the government's rule. For these reasons we will refer to ρ_i^* as i's "policy entitlement".

We will see later that disappointing any of these two moral expectations, the former towards the government and the latter towards others, will fuel i's feelings of distrust towards the government and towards others, respectively.

We can now derive ρ_i^*. Let

$$V_i(\rho, m) \equiv \varepsilon_i \left(b_i^*(\rho, m) + B_{-i}^*(\rho, m)\right) - c(b_i^*(\rho, m)) - \overline{k} \cdot p(b_i^*(\rho, m), \rho, m) \tag{8}$$

denote i's indirect utility function. Given the policies ρ and m, all individuals choose their equilibrium behavior as specified by Lemma 1. The "moral expectation" that all individuals will abide by the rule is formally equivalent to assuming that $m = 1$ (see expression (7)). Thus, individual i's policy entitlement is the solution of the following problem:

$$\rho_i^* = \arg\max_{\rho} V_i(\rho, 1) \tag{9}$$

In this ideal world, everyone complies. Society is thus split in two groups (cf. Lemma 1.4). The first group consists of the *affected* individuals, who mitigate their behavior to *fully* comply with their own rule ρ_i^*. These are individuals who, because of the rule, have to increase their behavior above their decentralized equilibrium behavior, b_i^* as defined by (3), and pay an additional adjustment cost.

The second group consists of the *unaffected* individuals. Once their own preferred rule is set up, their behavior is (weakly) higher than the rule. These individuals save

on adjustment costs with respect to the decentralized equilibrium because, as soon as the rule forces the first group to increase their behavior, they can decrease theirs due to strategic substitutability.

The optimization problem defined by (9) depends on whether individual i is affected or not by her optimal rule. If she is affected, her behavior is exactly ρ, and her optimality condition is the following:

$$\varepsilon_i'(\rho + B_{-i}^*(\rho, 1)) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho, 1)}{\partial \rho}\right) = c'(\rho). \tag{10}$$

By (10), i equates the marginal benefits of increasing her own behavior and the average behavior of others against the marginal cost due to complying. She also accounts for non-affected individuals reducing their behavior in equilibrium, due to strategic substitutability. As a result, she chooses a rule, ρ_i^*, that is higher than her decentralized behavior with no rule, b_i^*. The reason is simple: the prospect of pushing low types to increase their behavior and benefit from their externality leads i to choose a rule that is stricter than her decentralized behavior, despite paying an additional adjustment cost.[18] In sum: $b_i^* < \rho_i^* = b_i^*(\rho_i^*, 1)$.

If i is in the second group, i.e. the group of unaffected individuals, her equilibrium

[18] More precisely, the total behavior of other individuals is

$$B_{-i}^* \equiv \int_0^{x(\rho,1)} \rho dj + \int_{x(\rho,1)}^1 b_j^*(\rho, 1) dj$$

thus, the LHS of (10) can be written as

$$\varepsilon_i'(\rho + B_{-i}^*(\rho, 1)) \cdot \left(1 + x(\rho, 1) + \int_{x(\rho,1)}^1 \frac{\partial b_j^*(\rho, 1)}{\partial \rho} dj\right)$$

The first bracketed term accounts for the effect of the rule on i's behavior. The second one is the (positive) effect on all the other affected individuals up to type $x(\rho, 1)$, and the third is the (negative) effect on the non-affected individuals who will in fact reduce their behavior due to strategic substitutability. Observe that they will decrease their behavior as long as $\frac{\partial B_{-i}^*(\rho,1)}{\partial \rho} > 0$, which in turn implies that $x(\rho, 1) + \int_{x(\rho,1)}^1 \frac{\partial b_j^*(\rho,1)}{\partial \rho} dj = \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho} > 0$. The second bracketed term in (10) is thus positive, which in turn implies that the rule satisfying (10) must be higher than the decentralized behavior satisfying (3).

The second-order condition is

$$\varepsilon_i''(\rho_i^* + B_{-i}^*(\rho_i^*, 1)) \left(1 + x(\rho_i^*, 1) + \int_{x(.)}^1 \frac{\partial b_j^*}{\partial \rho} dj\right)^2 + \varepsilon_i'(\rho_i^* + B_{-i}^*(\rho_i^*, 1)) \cdot \left(x'(.) + \int_{x(.)}^1 \frac{\partial^2 b_j^*}{\partial \rho^2} dj\right) < c''(\rho_i^*, 1).$$

It is satisfied with a sufficient degree of cost convexity and benefit concavity.

behavior, $b_i^*(\rho,1) \geq \rho$, is not directly affected by the rule and her optimality condition is

$$\varepsilon_i'\left(b_i^*(\rho,1)+B_{-i}^*(\rho,1)\right)\cdot\left(\frac{\partial b_i^*(\rho,1)}{\partial\rho}+\frac{\partial B_{-i}^*(\rho,1)}{\partial\rho}\right)\geq c'(b_i^*(\rho,1))\frac{\partial b_i^*(\rho,1)}{\partial\rho}. \tag{11}$$

If the rule is interior with (11) holding with equality, i chooses a rule that is strictly lower than her equilibrium behavior given the rule, $b_i^*(\rho,1)$. In this case benefits are twofold. First, she can benefit from more mitigating behavior by others, on average. Second, she can lower her behavior (by strategic substitutability) and thus save on adjustment costs, which is an additional benefit. The cost lies rather in the fact that reducing own behavior lowers her benefits directly (because $\frac{\partial b_i^*(\rho,1)}{\partial\rho}<0$). The rule ρ_i^* equates benefits and costs at the margin. Note that i's behavior is also lower than her decentralized behavior, $\rho_i^* < b_i^*(\rho_i^*,1) < b_i^*$.[19]

When (11) is a strict inequality, ρ_i^* is equal to i's equilibrium behavior given the rule. In this case, the prospect of saving on adjustment cost is not that attractive compared to benefits. Thus i chooses the highest possible rule, which is exactly equal to her equilibrium behavior given that rule, $b_i^*(\rho_i^*,1)$. Intuitively, this configuration is more likely to happen if costs are low at the margin, and thus if i is not very concerned. By strategic substitutability, i's behavior is lower than decentralized behavior with no rule, $\rho_i^* = b_i^*(\rho_i^*,1) < b_i^*$.

By implicitly differentiating (10) or (11), it follows that policy entitlements are (weakly) monotonic in individuals' types, $\frac{\partial\rho_i^*}{\partial i}\geq 0$. More concerned individuals also believe that the rule should be (weakly) more stringent.

2.3 The Government Policy

Similar to individuals, also the government has got its own view about the true benefits of mitigating behavior, and paternalistically uses that view to set up what it deems the right policy. We do not take a specific stance about the government's degree of concern. It can be high or low. To simplify our discussion, we assume it lies somehow in between the lowest and the highest degree of concern in society, $\varepsilon_0'(.) < \varepsilon_G'(.) < \varepsilon_1'(.)$. We moreover remain agnostic about where the government's view comes from. Of course it

[19] To illustrate this result, suppose I am quite concerned about Covid-19. With no rule, my typical (quite high) social distance is $b_i^* = 3$ yards because I am afraid of some people who keep no distance at all. If I could set up a $\rho_i^* = 1$-yard minimum rule for everyone, then I would be less afraid and lower my social distance to just $b_i^*(\rho_i^*,1) = 2$-yards.

is important to understand what shapes it. It may derive from specific information, or by the composition of contrasting views of parties, lobbies or pressure groups, or simply reflect the average sentiment amongst citizens. Here we assume for simplicity that $\varepsilon_G(.)$ is exogenously given as we want to focus on the citizens' side and explore how their heterogeneous views shape their feelings of distrust.[20]

Unlike individuals, who compute their preferred rule by thinking of an ideal world in which everyone complies, the government faces the enforcement problem pragmatically. Therefore, along with its rule denoted by ρ_G^*, it simultaneously chooses the optimal level of monitoring, m_G^*. The policy problem is then:

$$\rho_G^*, m_G^* = \arg\max_{\rho,m} W(\rho, m)$$

$W(\rho, m)$ is a paternalistic indirect social welfare function built under the assumption that any citizen is assigned what the government deems the true benefit function, $\varepsilon_G(.)$:

$$W(\rho, m) \equiv \int_0^1 \varepsilon_G\left(b_i^*(\rho, m) + B_{-i}^*(\rho, m)\right) - c(b_i^*(\rho, m))di - k(m) \tag{12}$$

with $k(m)$ being the cost of monitoring individuals' behavior, which we assume is an increasing, convex, and twice differentiable function. To get rid, for simplicity, of re-distributive and double-dividend concerns, here we also assume that fines proceeds are burnt. The optimal behavior of citizens, $b_i^*(\rho, m)$, is the one specified by Lemma 1. The government therefore correctly anticipates the reaction of every citizen to policies ρ and m, which in fact is based on their individual perceptions of the externality, $\varepsilon_i(.)$. In other words, the government chooses the policy under the presumption that all individuals will actually benefit from their behavior according to $\varepsilon_G(.)$, but is perfectly aware that the citizens will react to that policy based on their own idiosyncratic evaluations of the benefits of mitigating their behaviors.

Because enforcement might not be perfect, some individuals might choose not to comply. In this case, society would be split in three groups (cf. Figure 1). The first group are the affected individuals choosing not to fully comply. These are the least

[20]This model could be extended to include elections, lobbies, protests, and other forms of political pressure that might affect government's views or pull it away from paternalitic attitudes. Such extensions would yield qualitatively similar predictions as the present model, as long as the government's concern lies somewhere between the least and the most concerned type. Proposition 3 shows how exogenous changes in the government's view and policy affect individuals' trusting attitudes

worried with $i \in [0, z(\rho, m))$. The second group are the affected individuals choosing to fully comply. They are relatively less worried, with $i \in [z(\rho, m), x(\rho, m)]$. The third group are the unaffected ones whose mitigating behavior is higher than the government rule. They are most concerned, with $i \in (x(\rho, m), 1]$. Hence, the government's objective function becomes,

$$W(\rho, m) = \int_0^{z(\rho,m)} \left[\varepsilon_G\left(b_i^*(.) + B_{-i}^*(.)\right) - c(b_i^*(.))\right] di$$
$$+ \int_{z(\rho,m)}^{x(\rho,m)} \left[\varepsilon_G\left(\rho + B_{-i}^*(.)\right) - c(\rho)\right] di + \int_{x(\rho,m)}^{1} \left[\varepsilon_G\left(b_i^*(.) + B_{-i}^*(.)\right) - c(b_i^*(.))\right] di - k(m),$$

The three integrals are aggregate utilities of the three groups. Let us assume for simplicity that the government's optimal rule and monitoring are interior.[21] The associated first-order conditions are: $\frac{\partial W}{\partial \rho} = 0$ and $\frac{\partial W}{\partial m} = 0$, which can be written as

$$\int_0^{z(.)} \varepsilon_G'(b_i^*(.) + B_{-i}^*(.)) \left(\frac{\partial b_i^*}{\partial \rho} + \frac{\partial B_{-i}^*}{\partial \rho}\right) di + \int_{z(.)}^{x(.)} \varepsilon_G'\left(\rho + B_{-i}^*(.)\right)(1 + \frac{\partial B_{-i}^*}{\partial \rho}) di$$
$$+ \int_{x(.)}^{1} \varepsilon_G'\left(b_i^*(.) + B_{-i}^*(.)\right) \left(\frac{\partial b_i^*}{\partial \rho} + \frac{\partial B_{-i}^*}{\partial \rho}\right) di = \int_0^{z(.)} c'(b_i^*(.)) \frac{\partial b_i^*}{\partial \rho} di + \int_{z(.)}^{x(.)} c'(\rho)) di + \int_{x(.)}^{1} c'(b_i^*(.)) \frac{\partial b_i^*}{\partial \rho} di \tag{13}$$

and

$$\int_0^{z(.)} \varepsilon_G'(b_i^*(.) + B_{-i}^*(.)) \left(\frac{\partial b_i^*}{\partial m} + \frac{\partial B_{-i}^*}{\partial m}\right) di + \int_{z(.)}^{x(.)} \varepsilon_G'\left(\rho + B_{-i}^*(.)\right) \frac{\partial B_{-i}^*}{\partial m} di$$
$$+ \int_{x(.)}^{1} \varepsilon_G'\left(b_i^*(.) + B_{-i}^*(.)\right) \left(\frac{\partial b_i^*}{\partial m} + \frac{\partial B_{-i}^*}{\partial m}\right) di = \int_0^{z(.)} c'(b_i^*(.)) \frac{\partial b_i^*}{\partial m} di + \int_{x(.)}^{1} c'(b_i^*(.)) \frac{\partial b_i^*}{\partial m} di + k'(m) \tag{14}$$

The three integrals on the LHS of both expressions are the aggregate marginal benefits of individuals' behavior in the three groups. They can be either positive or negative (see below). The government's policies also affect the size of the three groups. But these size effects are offset at the margin. The government's optimal policy mix simultaneously solves the two FOCs above.

By (13), the choice of the optimal rule implies several tradeoffs. A higher rule yields *more* mitigating behavior by those who fully comply (in the second group - second integral on the LHS). Perhaps less intuitively, it also yields *less* mitigating behavior by the non-compliant ones in the first group (first integral), because the marginal benefit of mitigating behavior is smaller when the rule is stricter (cf. Lemma 1.3). Moreover, a stricter rule

[21] Nothing crucial hinges on this assumption, except that the optimality conditions are different when $m_G^* = 1$, becausem by Lemma 1, society is split in two groups instead of three.

leads the very concerned types in the third group to *lower* their behavior due to strategic substitutability (third integral). Still the overall behavior in society increases with the rule, because otherwise individuals in the first and the third group would not lower their behavior (and we would get a contradiction). The government also takes into account aggregate adjustment costs at the margin.

There are also several trade-offs when choosing monitoring activity. Because stricter monitoring increases the risk of being caught, non-compliant individuals in the first group increase their behavior, yielding a positive effect at the margin (first integral in (14)). Stricter monitoring yields higher overall-average behavior, which positively affects the benefits in the second group (second integral). The effect on the third group's benefits is negative because they lower their behavior due to strategic substitutability (third integral). The government also takes the marginal cost of monitoring and citizens' marginal adjustment costs into account.

An increase in the government's level of concern (a higher ε'_G) shapes the policy mix. In general, the government chooses to increase both the rule and monitoring. However, raising the rule and intensifying monitoring have contrasting effects on the non-compliant group. If the benefit function is quite concave and the rule is already strict, the government might find it convenient to lower the rule and increase monitoring in order to foster mitigating behavior of the non-compliant citizens (whose mitigating behavior is more valuable at the margin). This would come at the cost of less mitigating behavior of the second group (and eventually the third one), which in fact is less valuable at the margin. This model suggests that when the benefits of mitigating behavior occur primarily at low levels of behavior, government will tend to rely more on monitoring than on regulation. The more so, the more the government is concerned.

2.4 Institutional and Interpersonal Distrust

As in Coleman (1994) and Hardin (2002, 2006), our conception of trust is grounded in individual expectations of trustworthy behavior: the trustor will trust the trustee if she expects the latter to act in her interest. In the case of institutional trust any individual i then becomes the trustor, the trustee is the government, and the expected behavior is implementing ρ_i^*, which is the "right" policy in individual i's view. If the government fails to do so, she loses trust in political institutions. Conversely, interpersonal trust follows

from the trustworthy behavior of fellow citizens, which now reflects a 'civic' expectation of respecting the rule ρ_G^* set by the government. If others do not respect the rule, individual i will start distrusting them as well.

In this formulation, political and interpersonal trust are two complementary elements of an individual's moral view on society. First, a government is trustworthy when it implements policies reflecting her view of how society should be governed. Second, citizens are trustworthy when they do what every civic citizen should do in a democracy: abide by the government's rules.[22]

In what follows we describe in detail how we operationalize both notions of trust in the model and how our definitions of social and political trust can be reconciled with existing definitions in the literature.

2.4.1 Institutional Distrust

An individual i feels that the "right" policy is ρ_i^*. This policy reflects her genuine perception of the externality problem. Any other policy would raise individual i's concern that the government is not acting in the citizens' interests. Reasonably, this concern increases in the loss caused by the government's actual policy ρ_G^* relative to ρ_i^*. We thus define institutional distrust as follows,

Definition 1. *Let D_i^I denote i's institutional distrust. We assume it is given by the utility loss due to adopting ρ_G^* instead of ρ_i^*. We then obtain for all i,*

$$D_i^I(\rho_G^*) = V_i(\rho_i^*, 1) - V_i(\rho_G^*, 1) \tag{15}$$

$V_i(\rho_i^*, 1)$ is the counterfactual utility that i would enjoy if the government policy was ρ_i^*, and $V_i(\rho_G^*, 1)$ is a level utility that she actually enjoys under the government rule ρ_G^*.

Importantly, both utility levels are computed under the counterfactual expectation that monitoring is perfect so that $m{=}1$. Hence all individuals, including herself, will obey the rule. Under this assumption citizens' behavior would be trustworthy, while the government's behavior is not. Thus the loss of utility is entirely due to the government. The normative valence of ρ_G^* and $m{=}1$ derives from the fact that $V_i(\rho_i^*, 1)$ is not only

[22] This experience-based concept of trust aligns well with Levi and Stoker (2000). The same goes for the trust definitions by Delhey and Newton (2005), as "the belief that others will not deliberately or knowingly do us harm, if they can avoid it, and will look after our interests, if this is possible".

the utility i feels entitled to, but also the utility level she thinks every individual should receive in her view of a well-governed society.

Our conceptualization of institutional distrust aligns with the literature in several ways. Trust in public institutions tends to form an underlying dimension with two overlapping clusters within it. First, trust in the more neutral and impartial institutions of the state (the courts, police, and civil service) and, second, trust in the organizations of government in the forms of cabinet, parliaments, and political parties. It is the latter that we focus on here. However, as noted by Newton et al. (2018), the two dimensions of institutional trust tend to go together.

Gecewicz and Rainie (2019) have derived the expected trustworthy behavior of political institutions from the degree to which they embody four attributes: efficiency, care/empathy/benevolence, transparency, and accountability. These are widely used to measure institutional trust in surveys (e.g. the World Value Survey or the European Social Survey). Of these four attributes, efficiency is the one primarily captured by (15), based on individual views over the size of the utility loss suffered when the government fails to deliver what they think is the right policy. When utility falls short of the reference level, $V_i(\rho_i^*, 1)$, an individual might think public resources are used in an irresponsible and inefficient way. The utility loss can also be seen as the disappointment of not being heard or cared for by the government, which links our definition to the attribute of care/empathy/benevolence.[23]

One might observe that our definition of institutional trust in (15) - and possibly the definition of interpersonal trust below - requires a certain degree of sophistication by individuals. They are supposed to be able to see the counterfactual and to compute utility deriving from it. However, the predictions of this model would not change substantially if one assumed that individuals are naive and that $V_i(\rho_i^*, 1)$ merely reflects some sort of self-serving view of the level of utility that the government should provide. On the other hand, one could also observe that our definition of institutional distrust is perhaps too simple to account for situations in which the government is subject to certain political constraints. For example, individual i might understand that ρ_G^* is the only feasible policy given those constraints, or it is the policy reflecting the will of the majority of

[23]Lastly, transparency and accountability lie beyond the scope of our interest here because our individuals are assumed to have perfect information. Moreover, we have abstracted from the modeling of elections or other forms of political representation.

which she does not feel a part. In such cases, she would possibly shift the blame to other agents. These aspects could be captured by explicitly introducing political constraints into the government maximization problem. We do not do this here for simplicity, but doing it would be an interesting extension of our approach, possibly leading to a more articulated and multifaceted notion of institutional trust. Nevertheless, we believe it would not substantially change the main message of this paper.

2.4.2 Interpersonal Distrust

Adherence to civic or cooperative norms can mitigate complex coordination failures in a densely interwoven society (Guiso et al., 2011), such as the externality problem we model here, and can therefore be seen as an indicator of trustworthy behavior. Delhey et al. (2011), using "adherence to civic norms" as one of the key civic attitudes taken from the World Value Survey, indeed find a strong relationship between a society marked by norm adherence and the level of expressed interpersonal trust. In Tabellini (2010) the notion of social trust is also shown to have a strong correlation with generalized morality, measured by the value attached to "respecting other people" as a fundamental belief that should be transmitted from parents to children, as well as by the importance of obedience.

For this reason, the government rule ρ_G^* in our model shapes an individual's *moral* expectations of trustworthy behavior. Once in place, the policy serves as a reference point for evaluating the behavior of others, establishing a standard of civic responsibility and hence, trustworthiness. As a result, if an individual realizes that other people do not adhere to the rule, she will distrust them. This conceptualization is also in line with the view that social trust derives from a "belief that others will not deliberately or knowingly do us harm", as defined by Delhey and Newton (2005).[24]

Specifically, we assume that interpersonal distrust is expressed by the loss of utility caused by low compliance, compared to the counterfactual utility in which everyone complies. We rule out moral hypocrisy by assuming that when an individual is not (or only partially) following the rules, she can only distrust people whose behavior is even worse. Hence,

Definition 2. *Social distrust, D_i^S, is the loss of utility due to low compliance. For all i*

[24] A view that is also grounded in the work of Rousseau et al. (1998), which emphasizes the importance of interdependence, risk, and vulnerability in the relationship between trustor and trustee.

it is commensurate to the difference between i's reference utility and her actual utility in equilibrium, so that

$$D_i^S(\rho_G^*, m_G^*) = \begin{cases} V_i(\rho_G^*, 1) - V_i(\rho_G^*, m_G^*) & \text{if} \quad b_i^* \geq \rho_G^* \\ V_i(\rho_G^*, m_G^*)|_{b_j^* \geq b_i^*} - V_i(\rho_G^*, m_G^*) & \text{if} \quad b_i^* < \rho_G^* \end{cases} \tag{16}$$

The first line of (16) refers to compliant individuals. $V_i(\rho_G^*, 1)$ is the counterfactual utility that i would enjoy if all others behaved civically, while $V_i(\rho_G^*, m_G^*)$ is her actual utility. The second line of (16) accounts for the distrust of a non-compliant individual. $V_i(\rho_G^*, m_G^*)|_{b_j^* \geq b_i^*}$ is the counterfactual utility i would enjoy if all others mitigated their behavior at least to the extent that she does. Thus her loss of utility can only result from those behaving worse.

Ours is a notion of "generalized" trust as opposed to "particularized" trust (Putnam et al., 2000). The loss of utility is due to the non-compliant behavior of unknown, or socially distant people. Thus, it is linked to Yamagishi (2011) view that generalized trust is the default expectation of another person's trustworthiness "in the absence of information about that person."

In our formulation, institutional and social trust are complementary, and yet conceptually distinct. Individual i will not blame the government for the uncivil behavior of non-complying individuals, but the individuals themselves. On the other hand, she will not blame other individuals if the government does not implement what she thinks is the right policy.[25]

Of course, in the context of our model, one might think of different definitions of social and institutional trust, but at the risk of losing consistency. For instance, if an individual distrusts others just because their behavior does not match her entitled policy, she would blame others for the same reason why she blames the government. Inversely, if an individual were to distrust others for following a rule that she believes is wrong, a

[25]While it might be the case, in (16) we are assuming for simplicity that citizens do not blame the government for failing to fully enforce the chosen rule ρ_G^*. They understand monitoring is costly and think m_G^* is correctly determined. Not much hinges on this assumption. Predictions of the model would be similar if one assumes that not only has individual i an expectation about the rule, but also about the monitoring activity, say m_i^*. In this case, an individual would distrust the government for not implementing both the right rule and the right enforcement policy. She would place some of the blame for citizens' bad behavior on the government, while being more lenient with those individuals who do not strictly follow the rule.

similar issue would arise.

Finally, it is worth reiterating that our notions of social and institutional trust are derived from counterfactual hypotheses that reflect a moral vision of society, in which the government takes care of the citizens, and each citizen is called to do his or her part. Thus, each citizen expects the government to implement what she thinks are the right policies, and others are expected to behave ethically, where ethical behavior consists of respecting the rules, and respecting the rules is also a way of respecting others (Tabellini (2010)). We could have imagined a world in which an individual might sympathize with those who disobey the rules because they have a different view of the externality, and thus the harm caused by their misbehavior is somehow excused. But we would not have captured the ethical aspects mentioned above.

2.5 Equilibrium

We are now ready to define and characterize the political equilibrium.

Definition 3. *A political equilibrium consists of:*

i. A set of entitled rules, ρ_i^*, maximizing individuals' indirect utilities defined by (9) and taking into account the equilibrium behavior of the other individuals, $b_j^*(\rho_i^*, 1)$, under the assumption that everyone abides by ρ_i^*.

ii. A policy vector (ρ_G^*, m_G^*) maximizing the government's paternalistic welfare function defined by (12), taking into account individuals' equilibrium behavioral responses, $b_i^*(\rho_G^*, m_G^*)$.

iii. A set of individual equilibrium behavioral responses, $b_i^*(\rho_G^*, m_G^*)$ defined by Lemma 1, such that society is split in three groups as follows: 1. a group of unaffected individuals with $b_i^*(.) > \rho_G^*$ and $i > x(.)$; 2. a group of affected individuals deciding to comply with $b_i^*(.) = \rho_G^*$, and $i \in [z(.), x(.)]$; 3. a group of affected individuals deciding not to comply, with $b_i^*(.) < \rho_G^*$ and $i \in [0, z(.))$.

iv. A set of individual reference utilities, $V_i(\rho_i^*, 1)$, based on the presumption that all individuals comply with the entitled policy, ρ_i^*; two sets of counterfactual utilities, $V_i(\rho_G^*, 1)$ and $V_i(\rho_G^*, m_G^*)|_{b_j^* \geq b_i^*}$, based on the presumption that all individuals comply

with the government rule ρ_G^* *or, if* $b_i^*(\rho_G^*, m_G^*) < \rho_G^*$*, that all other individuals mitigate at least as much as individual* i *does.*

v. A set of equilibrium levels of institutional distrust, $D_i^I(\rho_G^*)$*, defined for all* i *by (15), and a set of equilibrium levels of interpersonal distrust,* $D_i^S(\rho_G^*, m_G^*)$*, defined for all* i *by (16).*

This equilibrium features a society where, roughly speaking, the government paternalistically implements a rule reflecting its own level of concern. Monitoring is costly, so it is not perfect. Some low-concern individuals choose not to fully comply. Some high-concern individuals display even higher mitigating behavior than prescribed by the rule. Virtually any citizen, to some extent, distrusts the government for not implementing what she deems the "right" rule, and any citizen distrusts other people for not complying with the government rule, or for complying less than she does.

In this society institutional distrust is the result of divergent responses to government policies. On the one hand, individuals with a high degree of concern have a strong appreciation for the benefits of mitigating behavior. As a result, they hold the government responsible for what they perceive as an insufficiently strict rule that fails to fully capture these benefits. However, their sense of disappointment is partly dampened by two factors. First, the missed benefits are decreasing at the margin. Second, by adjusting their behavior to the (lower) government rule, they incur lower mitigation costs than under their preferred rule. Intuitively, this dampening effect can be significant in scenarios where costs are increasing and benefits are decreasing rapidly at the margin, resulting in a relatively low level of disappointment in the high-concern group.

Conversely, low-concern individuals blame the government for bearing what they perceive as exorbitant adjustment costs. The more so, the more the costs are convex. The higher benefits associated with the government's rule serve to dampen their disappointment. But this dampening effect is small when the benefits are quite concave, leading to an overall high degree of disappointment in the low-concern group.

Thus, the model predicts an interesting asymmetry in institutional distrust. When complying with a regulation that imposes "sufficiently" convex costs and provides "sufficiently" concave benefits, the low-concern group will distrust the government more than the high-concern group. This result is illustrated by the left panel of Figure 2 and formally stated by Proposition 1-*i*. The proposition also states that the reverse is true.

In fact, with low cost-convexity and benefit-concavity, high-concern people will blame the government more because the rule fails to capture the benefits of mitigation (which are large for them at the margin). Low-concern people will blame the government less because the (useless) compliance costs are small at the margin.

At the risk of oversimplification, the general idea is that when mitigating behavior imposes significant costs, people with a lower sensitivity to externalities will complain more about the government. Conversely, if those with the highest levels of concern believe that the externalities in need of correction are particularly severe at the margin, they will express greater distrust of government.[26]

Let us now look at interpersonal distrust. In the view of a highly concerned individual, the misbehavior of those who disobey the rule causes large losses relative to a fair counterfactual in which everyone civically obeys the rule. Thus, we expect high-concern individuals to develop strong feelings of interpersonal distrust. In contrast, a low-concern individual does not view the misbehavior of others as a great loss. Thus her disappointment will be small. Moreover, if she chooses not to comply she can only be disappointed by the misbehavior of those who are less compliant than she is.

In sum, compared to high-concern individuals, the low-concern individuals are expected to develop a lower sense of interpersonal distrust. This result is stated in Proposition 1-*ii* and presented graphically in the right panel of Figure 2.[27]

Proposition 1 (**Asymmetry in distrust**). *Let $g \in [0,1]$ be the individual with a moderate level of concern, whose entitled policy is equal to the government's policy, $\rho_g^* = \rho_G^*$, and let*

- l denote any low-concern individuals such that $\rho_l^* < \rho_G^*$, and

- h denote any high-concern individuals such that $\rho_G^* < \rho_h^*$.

i. For any 'symmetric' couple of low-concern/high-concern individuals such that $|l - g| =$

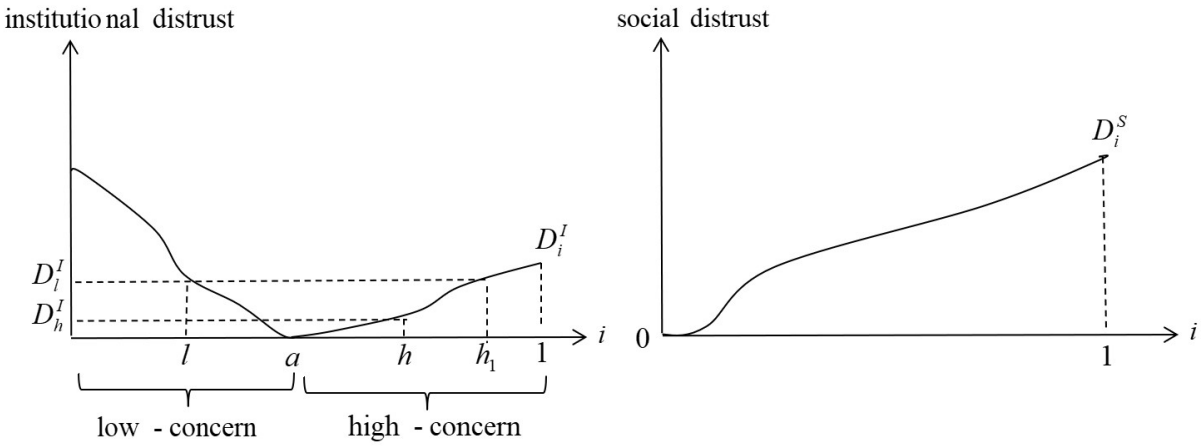
[26] It is important to remember that we have assumed that the government has a "moderate" level of concern. If we instead assume, for example, that the government's view is similar to that of the most concerned ones, then the least concerned group will exhibit greater levels of institutional distrust, regardless of the shape of the cost and benefit functions. The reverse is also true. We return on this in Proposition 3 below.

[27] Of course, moral hypocrisy might lead a noncompliant individual to blame people who are complying more than herself. This would be an interesting attitude to study, but, as pointed out earlier, we do not include it here. Another interesting factor affecting interpersonal distrust would be people's propensity to blame others for getting the government to adopt a policy they do not like - through voting, lobbying, or other political mechanisms. Since for simplicity in this paper we abstract away from any political distortions behind the government's choice, we do not consider these types of factors here. They are nevertheless interesting, but would probably require more specific investigation.

$|h - g|$, *high (resp. low) convexity in the cost function and high (resp. low) concavity in the externality functions lead the low-concern individual to distrust the government more (resp. less) than the high-concern individual.*

ii. For any m_G^**, the higher an individual's level of concern, the more that individual distrusts other people.*

Figure 2: Asymmetry in trust



While Proposition 1 points to the shape of the benefit and cost functions, another crucial factor playing a role in our theory is heterogeneity of beliefs. Since institutional distrust arises from disagreement with the government rule, we expect higher institutional distrust in societies with more diversified beliefs about the severity of the externality. Holding the government rule fixed, and assuming that it reflects the average level of concern in society, the more individuals disagree about the right policy to address an externality problem, the more distrust of the government will be observed.

In addition, more heterogeneity in beliefs will lead to larger differences in behavior and deviations from the government's rule. This in turn will lead to greater harm to the most concerned individuals, resulting in larger losses of interpersonal trust.

Finally, the effects of regulation on institutional and interpersonal distrust will be more pronounced in societies where the level of concern increases proportionally for all individuals.

These results suggest that, despite the government's best efforts to implement a rule that pleases the average citizen, distrust will still increase as the level of concern becomes proportionally higher or more broadly dispersed within society. Proposition 2 formalizes these results. In fact, it can be shown that the first statement of the proposition holds for any government rule, leading to an even more pessimistic view of how belief heterogeneity affects political and interpersonal trust.

Proposition 2 (**Disagreement and Concern**). *Let* $TD^I \equiv \int_0^1 D_i^I di$ *be total institutional distrust in a society, and let* $TD^S \equiv \int_0^1 D_i^S di$ *be total social distrust. Let a denote the citizen with the average level of concern, such that* $\epsilon'_a(y) \equiv \int_0^1 \epsilon'_i(y) di$*, and the identity, a, of the mean-concern individual is the same for any values of y. For any government policy reflecting the average citizen's level of concern,* $\rho^*_G = \rho^*_a$*,* TD^I *and* TD^S *increase if*

i. disagreement about the severity of the externality increases, where disagreement is captured by a mean-preserving spread of $\varepsilon'_i(y)$*;*

ii. the level of concern is sufficiently large and increases proportionally for all individuals.

Whereas Proposition 2 focuses on overall levels of trust, the following two propositions explore how a given shock affects trust asymmetrically for more or less concerned individuals. Those propositions will guide the discussion of our empirical results in Section 3. We are interested in two comparative statics: when the government rule or mitigation costs are "perceived" to increase. Here we talk about perceptions because our experiment treatments in Section 3 are aimed at increasing individuals' perceptions about different aspects of the Covid-19 crisis, such as the strictness of the rules or the cost of mitigating behavior.

Proposition 3 assumes the government rule increases, holding monitoring equal. High-concerned individuals will like that policy more, while low-concerned ones will like it less, with institutional distrust decreasing for the former and increasing for the latter, and the aggregate effect being higher distrust if the rule increases a lot. Furthermore, a stricter rule will lead all individuals to distrust each others more, because more individuals will not comply and take the risk of being caught. This result echoes that of Aghion et al. (2010), despite a different mechanism being at play here.

Proposition 3 (**Strict intervention**). *If the government rule ρ_G^* is (perceived to be) increasing then, all else being equal,*

i. high-concern individuals distrust institutions less and low-concern distrust them more; the total effect on institutional distrust TD^I is positive if the perceived increase in ρ_G^* is sufficiently large, and,

ii. all individuals distrust others more; total social distrust TD^S increases; the effect is larger for high-concerned individuals compared to low-concerned ones.

The following proposition examines how the perceptions of mitigation costs affect trust. If costs are perceived to be higher by all individuals, they will prefer less restrictive rules. Given the government's rule, the highly concerned will agree more with the government and gain institutional trust, while the less concerned will disagree more and lose institutional trust. More individuals will choose not to comply or to further reduce their behavior, feeding the distrust of others, especially those who are more concerned.

Proposition 4 (**Costly compliance**). *If marginal mitigation costs $c'(b_i)$ are (perceived to be) increasing for any value of b_i, then, holding the government rule constant and all other things being equal, ideal mitigating policies ρ_i^* are lower for all individuals, so that*

i. high-concern individuals distrust institutions less and low-concern individuals distrust them more; the total effect on institutional distrust TD^I is positive (resp. negative) if the perceived increase in $c'(b_i)$ is sufficiently large (resp. small), and,

ii. all individuals distrust others more; total social distrust TD^S increases; the effect is larger for high-concern individuals compared to low-concern ones.

3 Experimental Effects of the Covid-19 Crisis on Distrust

We now present a large online survey experiment we conducted in June 2020 and interpret its findings through the lens of our theoretical model.[28] Our primary goal was to make

[28]The survey is part of a larger project studying the relationships between Covid-19 crisis and individuals' attitudes towards a large set of socio-political issues. Here we consider a subset of survey questions regarding institutional and interpersonal trust along with attitudes towards redistributive policies. The large questionnaire also included questions on populism, values and globalization, which we do not take up in the analysis here but account for when correcting for multiple hypothesis testing in Appendix C.6.

several dimensions of the Covid-19 crisis more salient to investigate the effect on trust in institutions and trust in others.

The treatment (Covidfirst) presents a randomly selected group of respondents with a block of treatment questions on how they experienced the pandemic, *before* asking them a block of outcome questions eliciting trust and socio-political attitudes. The control group (Baseline) instead receives the outcome questions *before* the treatment questions. Table 1 summarises this reversed-order experimental design, and paints a visual picture of the survey flow for the two conditions.

Table 1: Experimental Design

Covidfirst	**Baseline**
Background information	Background information
Treatment questions Sub-conditions: Health Health & Economic Health & Conflict	Outcome questions Institutional trust Interpersonal trust
Outcome questions Institutional trust Interpersonal trust	Treatment questions Sub-conditions: Health Health & Economic Health & Conflict
Further background information	Further background information

In line with other papers using similar priming techniques, our presumption is that the Covid-19 crisis becomes more salient by fielding the treatment questions before the outcomes. Any changes in the answers of treated respondents to outcome questions, compared to non-treated respondents, can be attributed to thinking more intensely about the pandemic, and proxy the way trust is affected (see Stantcheva (2023) for an illustration of priming in survey experiments).

To investigate the mechanisms driving these salience effects we divide the treatment questions into three blocks:

1. **Health**: These questions tackle respondents' day-to-day experience of the epidemic, e.g. which of the containment measures they have adopted, whether they had Covid-

The survey was pre-registered as AEARCTR-0005952. Heterogeneous treatment effects were not included in the pre-analysis plan. See Appendix D for a link to the questionnaire and a complete list of the outcomes.

19 cases among their acquaintances and family members, whether they found the policy measures too strict or hard to bear, whether they were concerned for their health. This block makes respondents think more intensely about the pandemic as a public health crisis, and about the (appropriateness of the) containment policies enacted by their governments.

2. **Economic**: This block targets respondents' experience with the economic consequences of the epidemic. We asked for instance whether they lost their job or any future job opportunities, or whether people around them were suffering from economic distress. These questions are expected to trigger the respondents' awareness of the economic effects of the crisis, along with the policies to alleviate economic hardship.

3. **Conflict and Cooperation**: These are questions shaped along a rally-around-the-flag rhetoric. For instance, we asked respondents whether the crisis can be compared to a war waged against a hidden enemy, and to what degree it requires unity and cooperation to win it. This block makes respondents think intensely about the social cohesion needed to overcome a crisis, and how important it is that people comply with the rules to do so.

The outcome questions elicit respondents' institutional and interpersonal trust:

Institutional trust: We ask respondents how much they trust politicians, the national government, and the European Union. In order to check for accuracy, we introduce a real-stakes task, similar to DellaVigna et al. (2016) and Dechezleprêtre et al. (2022). It asks respondents to spend time (5 minutes) to read a pro-European integration speech and give their opinion for educational purposes. We expect this measure to be positively correlated with trust in the EU specifically, and institutions more generally.[29]

Interpersonal trust: We elicit interpersonal trust using the World Value Survey question that asks respondents whether they believe "one can never be careful enough in dealing with people (1), or [...] most people can be trusted". As discussed by Delhey et al. (2011), this question is mostly applicable to the notion of generalised trust we are interested in.

[29]For a description of this real-stakes measure and its challenges see Appendix B.1 and B.3.

Importantly, we use the three treatment questions blocks to implement three mirroring treatments in Covidfirst. In the first one, the respondents receive 'Health' questions only. The second treatment consists of 'Health' plus 'Economic' questions, and the third of 'Health' plus 'Conflict and Cooperation' questions. We denote these treatments by H, H+E, and H+C, respectively. We include the 'Health' questions in all treatments because the 'Economic' and the 'Conflict and Cooperation' questions could activate health-related aspects as well. To cleanly identify the impact of the Economic and the Conflict and Cooperation dimensions in the second and third treatments, we intentionally activate and fix health aspects across all treatments.[30] For this reason we collect data on individuals' experience of the pandemic in the health domain from the entire sample. Because these questions are not expected to be sensitive to the experimental variation, we are able to run full sample heterogeneity analysis with respect to those variables.

We then estimate the following statistical model via OLS regression

$$Y = \beta_0 + \beta_1 T + \beta_2 X + \beta_3 W + \beta_4 \kappa + \epsilon,$$

where Y is the answer to any outcome question, T is an indicator variable equal to zero in the Baseline condition (no treatment) and equal to 1, 2, and 3 in the H, H+E, and H+C sub-conditions. X and W are individual and regional controls, whilst κ denotes regional (NUTS-2) fixed effects.[31] To ease the interpretation of our results, all outcome variables have been standardised with respect to the outcomes in the Baseline condition. In some regressions we pool all three treatments together, and refer to this as the "Covidfirst" condition. Table C7 reports summary statistics.

The survey was fielded in Germany, Italy, the Netherlands and Spain in the first two weeks of June 2020, in the aftermath of severe lockdown measures adopted in all four countries. A random sample of approximately 2,000 adult individuals per country were contacted by the survey company *Respondi* to distribute the link to our Qualtrics survey.

[30] We provide respondents in the 'Health' sub-condition neutral filler questions to ensure they do not receive disproportionately fewer questions than respondents in the 'Economic' and 'Conflict and Cooperation' conditions.

[31] All regressions control for gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income, a dummy indicating the position of the question we use to get a behavioral (incentivised) measure of support for the European Union. We cluster the standard errors at the finest level available for each country. Specifically, NUTS-3 for Italy and Germany and NUTS-2 for Spain and the Netherlands. Our results are unchanged by clustering at different levels.

The sample was built to be close to representatives in each country.[32]

3.1 Asymmetry in Trust

We start by presenting empirical evidence for Proposition 1 which predicts that (i.a) the relationship between health concerns and institutional distrust is U-shaped; (i.b) highest institutional distrust are displayed by least concerned individuals; (ii) interpersonal distrust is increasing in concern. These patterns are outlined by Figure 2 in Section 2. Correlations between individuals' health concerns and several trust measures provide evidence with no need of relying on treatment effects. As part of the 'Health' treatment block, individuals are asked to rate their level of concern about the virus on a scale from 1 to 10. Table 2 divides the sample into five quintiles based on their level of concern. The omitted category is the group of individuals with the lowest level of concern (first quintile). Column (4) shows that respondents' level of interpersonal trust decreases as their self-reported health-concern increases. The more an individual is concerned, the more she distrusts others, as predicted by Proposition 1.ii. Additionally, the table indicates higher levels of institutional trust for all the four groups (so lower levels of distrust) compared to the first group. Individuals in the fifth quintile exhibit lower levels of trust in the EU, the national government, and in politicians, compared to the third and fourth quintile. The relationship between these measures of institutional trust and concern is therefore an inverted-U, and consequently the relationship between institutional *dis*trust and concern is *U-shaped*, as predicted by Proposition 1.i.a. Importantly, it is asymmetric: the least concerned individuals are those distrusting institutions the most, as stated by Proposition 1.i.b. Unfortunately we did not find evidence of a U-shaped pattern in the relationship between concern and our incentivized measure of trust in the EU (Column 5).

Proposition 2 states that greater heterogeneity in concern leads to increased distrust

[32]See Appendix B for further details about the survey flow and the data collection, and Appendix D for links to the local language and English surveys). Table C24 shows that the randomization strategy was effective: no sizeable differences in these variables can be found between individuals in the Baseline and Covidfirst groups. As further discussed in Appendix B.3, we address demand effects by leaving the respondents blind to experimental conditions other than the one they are participating in. We moreover address potential fatigue effects by randomising the position of demanding sections of our survey orthogonally to our treatments. Finally, Appendix C.6 reports the corrected p-values for multiple hypothesis testing for the analyses presented in the next section. These analyses confirm the validity of our results and the robustness of our design.

in society, both social and institutional. We were unable to identify any dimension (e.g., regions, or other characteristics) in which the sample of respondents displayed a sufficient amount of variation in health-concern heterogeneity. Therefore, our data cannot provide direct support for this proposition.[33]

Table 2: Empirical evidence for Proposition 1: Correlation of Health Concerns and Trust

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Health Concern 2^{nd} Quintile	0.204***	0.265***	0.229***	0.014	-0.015
	(0.035)	(0.032)	(0.029)	(0.030)	(0.028)
Health Concern 3^{rd} Quintile	0.331***	0.403***	0.386***	0.041	-0.019
	(0.039)	(0.041)	(0.035)	(0.038)	(0.035)
Health Concern 4^{th} Quintile	0.411***	0.431***	0.407***	-0.032	0.103***
	(0.040)	(0.043)	(0.039)	(0.047)	(0.037)
Health Concern 5^{th} Quintile	0.288***	0.256***	0.259***	-0.192***	0.176***
	(0.041)	(0.050)	(0.050)	(0.050)	(0.033)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.063	0.165	0.221	0.082	0.089
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Health Concerns': perceived Covid-19 health risks from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

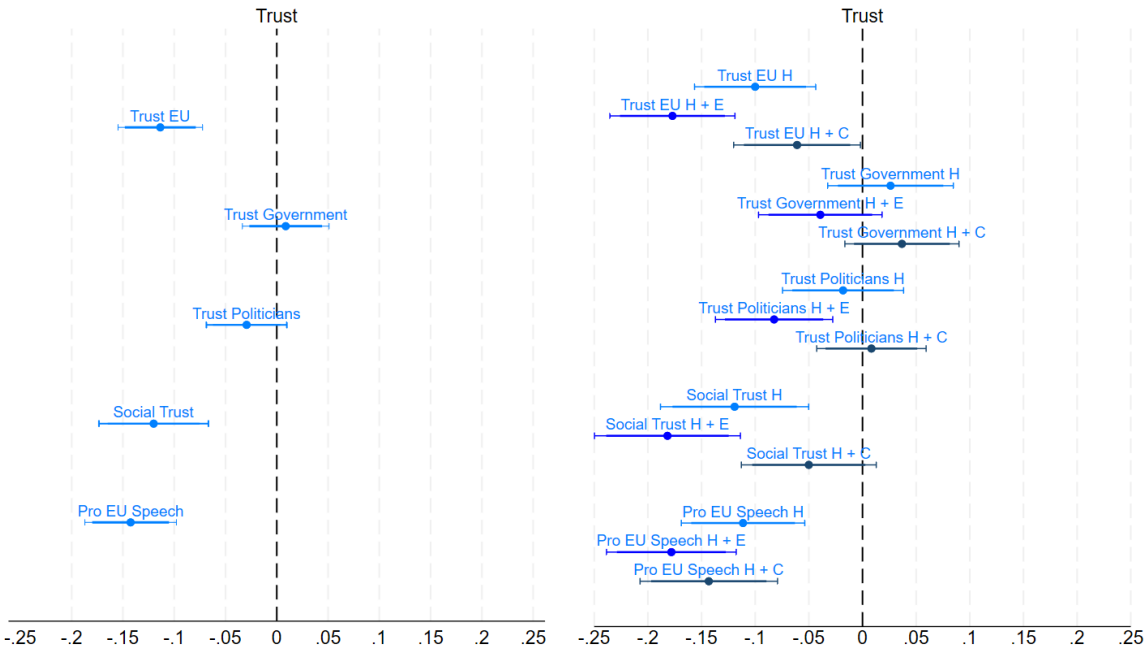
3.2 Treatment effects

The treatment was intended to lead participants to think intensely about the pandemic, the government rules, and the behavior of other people. We find that simply asking them questions about their pandemic experience before asking about trust makes them less trustful towards others and substantially affects their trust towards political institutions. This finding is illustrated in Figure 3. It shows the treatment effect on several trust dimensions and the associated confidence intervals, relative to the Baseline condition. In the left panel all three sub-conditions are pooled together. Following treatment, trust in the EU drops by -11.4% ($p < 0.01$) of the Baseline's standard deviation, and interpersonal trust drops -12% ($p < 0.01$). Primed respondents are 14% of a standard deviation less likely to read the Pro-EU speech. These results support the idea that when

[33]However, we can provide indirect evidence of Proposition 2, assuming that higher exposure to social media leads to greater heterogeneity in pandemic-related concerns. In Table C8, a dummy variable equal to 1 is coded for respondents who primarily receive information from social media, as opposed to traditional media such as TV news, newspapers, or radio/podcasts. The table shows a strong negative correlation between social media exposure and our five measures of trust. Assuming that increased exposure to polarized opinions on social media leads to greater heterogeneity in health concerns, this correlation supports Proposition 2.

thinking intensely about the crisis and about the government policies, individuals feel distrust more intensely. This is in line with our main theoretical point predicting that government intervention to curb externalities yields political and interpersonal distrust. The corresponding tables are reported in Appendix Tables C9 and C10.[34]

Figure 3: Treatment effects on trust



Left panel: all sub-conditions together. Right panel: separate sub-conditions Health (H), Health + Economic (H+E) and Health + Conflict/Cooperation (H+C). Confidence intervals: 95% (vertical bars) and 90% (bold line). Dependent Variables: **Social Trust**: on a scale from 1 to 10, do you think one can never be careful enough in dealing with people (1), or would you say that most people can be trusted (10)?; **Trust Government**, **Trust Politicians**, **the European Union**: on a scale from 0 to 10, how much do you trust each of the following: (1= not at all; 10= complete trust): Your national government; your national politicians; the European Union. See more details on the **Pro EU Speech** variable in Appendix B.

The treatment effect on trust in government and trust in politicians is non-significant. At first glance, this result might not seem consistent with our theory. However, our model offers two possible explanations.

First, the treatment may have reminded respondents in each of the four countries that the government rules were quite strict during the first Covid-19 wave in Spring 2020 (the survey was conducted immediately after the end of unprecedented lockdown policies). This might have led concerned individuals to think that the government is right in setting such restrictive rules, while it might have led unconcerned individuals to think that the government is wrong. If this is the case, then the treatment leads the former to trust the government more and the latter to trust it less. The two effects may offset each

[34]We ran all of our tests without covariates and obtain the same results in all cases. For the sake of brevity we do not report them here.

other resulting in the non-significant average effect observed in the data. Finding support for this explanation amounts to finding evidence for Proposition 3.*i*. It states that when the rule is stricter, high-concerned individuals will trust the government more while low-concerned individuals will trust the government less. Institutional distrust increases only if the effect on perceived rule is sufficiently high. Otherwise it might be ambiguous. Part 3.*ii* further states that interpersonal trust will decrease overall. Thus the average effect on institutional distrust may be ambiguous, while everyone would distrust others more. Intuitively, the effect would be a rightward shift of the curve in the left diagram of figure 3. and an upward shift of the curve in the right panel.

Table 3: Priming Covid-19, Trust and Health Concerns

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Covidfirst	-0.116***	0.007	-0.031	-0.119***	-0.143***
	(0.020)	(0.021)	(0.020)	(0.027)	(0.022)
Health Concerns	0.109***	0.067***	0.068***	-0.024	0.032**
	(0.023)	(0.022)	(0.022)	(0.022)	(0.015)
Covidfirst * Health Concerns	0.007	0.058**	0.060***	-0.041*	0.037*
	(0.029)	(0.025)	(0.023)	(0.024)	(0.019)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.058	0.156	0.214	0.079	0.087
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Health Concerns': perceived Covid-19 health risks from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3 provides support of Proposition 3 using the interaction between concern and the treatment. As expected, the first row reports no significant average effect of the treatment on trust in government and in politicians. However, once interacted with health-concern the treatment effect is significant and positive for highly concerned individuals in three out of four measures of institutional trust, including the incentivised outcome measuring political support for the European Union. The overall effect of the treatment on interpersonal trust is strongly significant. As expected, the interaction between concern and the treatment is negative when examining trust in others (column 4).

We replicate this analysis by dividing health concern in three tertiles. Results for this and the analyses discussed henceforth in this section are reported in Appendix C.3.

Table C4 shows that the negative treatment effects on trust in political institutions are driven by the least concerned individuals (first tertile), while the negative effects on interpersonal trust are driven by the most concerned ones (top tertile). This finding further corroborates Proposition 3.*i*, along with the idea that the treatment effects travel through the perception that rules were rather strict.[35]

Further evidence for these mechanisms is presented in Table C16 (Appendix C.4). There we run an analogous heterogeneity test using reported levels of compliance with government regulations. Consistent with our model, more compliant individuals lose trust in others while gaining trust in the government and politicians following our treatment.[36]

To delve deeper in the mechanisms behind the treatment effects, the right panel of Figure 3 breaks down the three sub-conditions. The non-significant effect of the treatment on trust in political institutions (left panel) is the composition of three mutually offsetting effects displayed in the right panel. In fact, trust in government and in politicians decreases (significantly) in the 'Economic' sub-condition (H+E) while it does not in the 'Conflict and Cooperation' (H+C) and 'Health' (H) sub-conditions. Moreover, the negative effect on social trust is stronger in the 'Economic' sub condition, compared to the other two treatments. It's possible that the three sub-conditions lead respondents to focus on different dimensions of the Covid-19 crisis with offsetting effects on trust, suggesting a second explanation for the overall insignificant effect of treatment on institutional trust. Specifically, while the 'Conflict and Cooperation' treatment (H+C) has focused individuals on the value of cooperation, the 'Economic' sub-condition (H+E) has focused them on the sizeable economic costs of the crisis, lockdown measures, and lack of compliance. Here we can find evidence of Proposition 4. It states that individuals prefer less severe mitigating measures when costs are higher, with the consequence that unconcerned individuals trust the government less, concerned individuals trust the government more, such that the total effect on institutional trust may be ambiguous (Proposition 4.*i*), while everyone distrusts others more (Proposition 4.*ii*).

Table C5 (Appendix C.3) provides support for this second explanation (and Propo-

[35]Table C16 in the Appendix reports similar findings when heterogeneity is captured by self-reported compliance with rules, measured by a specific question in the Health block. We find that respondents reporting higher compliance lose more trust in others while trusting the government and politicians more. This is not surprising since the theory predicts that people comply more if they are more concerned, and Table C6 indeed shows a strong correlation between compliance and health-concern (0.42).

[36]These results are robust to multiple hypothesis correction (Appendix C.6).

sition 4). It investigates the heterogeneous effects of health-concern along the three sub-conditions and shows that the interaction between concern and the 'Economic' sub-condition (H+E) on institutional trust is significant and positive (row 6, columns 2 and 3). Concerned respondents gain trust in the government once primed with economic dimensions of the crisis. The negative effect of the treatment on generalized trust is orthogonal to the three subconditions, as expected.

Table C5 shows that also the 'Conflict and Cooperation' sub-condition (H+C) yields a significant and positive effect on the institutional trust of concerned respondents (row 7, columns 2 and 3). The concerned turn to the government more than the low-concerned, once they are reminded of the importance of cooperation.

Table C6 (Appendix C.3) sheds further light on the mechanism connecting individuals' perceptions of the government's management of the pandemic with self-reported concern, compliance, and expectations of others' civic behavior. We included two questions on this matter: i) were the rules imposed by the government to face the pandemic too strict? ii) how well is the government dealing with the pandemic? Responses are strongly correlated with concern, compliance, and expectations about others' civicness. These correlations validate our theoretical predictions that more concerned individuals are less likely to think that rules are too strict and more likely to positively evaluate government's performance. The same applies to individuals complying more or expecting more civicness from others. In the 'Conflict and Cooperation' sub-condition we ask several questions on the degree to which respondents agree that cooperation and compliance with the rules is a way to defeat the virus. We interpret respondents' answers as their expectations about the civic behavior of others. We name the resulting factor variable 'Civic Expectations'. Our findings indicate a strong correlation between this variable and health concern (0.37) or self-reported compliance (0.48).[37] Individuals expecting civil behavior from others are likely to report health concern and compliance with Covid-19 rules. Somehow importantly, when using these expectations to capture heterogeneity, the treatment yields a sharp decrease in interpersonal trust among those expecting more civic behavior. This

[37]The variable 'Civic Expectations' is a combination of treatment questions resulting from a factor analysis. Specifically, we measure from 1 to 10 agreement with the statement that "We can defeat Covid-19 only if everyone self-sacrifices, e.g. by strictly respecting self-isolation at home"; "People breaking the rules can be considered traitors and should be punished"; "Unity is the main strategy to defeat the Covid-19 crisis". Logically, this is tested only on the subset of respondents in the 'Conflict and Cooperation' treatment (H+C) receiving this set of questions.

result supports our basic theoretical idea that social distrust is caused by disappointed expectations about the civic behavior of others (cf. Table C17 in the Appendix). It is also consistent with our non-hypocrisy assumption: those complying less also expect less from others.

4 Discussion: General validity and limitations

In Section 3 we show that our experimental effects align with the model's predictions. However, the priming effects observed in the survey experiment are likely to dissipate rapidly. Nevertheless, we contend that these experiments have empirical validity, as individuals are consistently exposed to on-the-field priming treatments in reality. They are continuously reminded of Covid-19, and these reminders serve as repeated priming experiments, prompting individuals to reflect on government policies and the behavior of others, which may perpetuate a loss of trust.

Our theory suggests that government regulation in response to new state of the world can lead to a trust crisis in society. However, the theory does not address the duration of this crisis. We can argue that if new trust crises occur before society has fully recovered from previous ones, trust will continue to decline over time. The theory also suggests a second reason why trust may erode over time: opinion polarization in societies. If individuals hold diverging beliefs on how an issue should be regulated, both trust in political institutions and in others would decline over time. This second reason points to the importance of information and the scientific methods to investigate the true nature of an externality problem.

Propositions 1 to 4 are consistent with experimental evidence regarding Covid-19. However, they can be given a general interpretation. They suggest that the risk of institutional and interpersonal distrust is latent in complex societies, where social coordination becomes more important, and individuals are strongly opinionized, if not polarized, in their beliefs. For this reason they can help understand what to expect from policies addressing climate change in the years to come.

Tackling climate change will indeed require enormous government intervention and restrictions on individual behavior. However, people disagree on the severity of the climate crisis. Many do not trust climatologists and are skeptical of the real risks of a

climate catastrophe (Algan et al. (2021)). And some in favor of restrictions today may not realize ex-ante how radically their behavior will have to change tomorrow.

Opinion surveys unambiguously show that heterogeneity of opinions regarding climate change does exist and may become more pronounced in the future. Some people think that we are already in a dramatic situation and that the government should act immediately. Others think the crisis has little to do with human behavior or at least it is not a top-priority. Both groups will be disappointed with the government's environmental policy. The former because it is doing too little, and the latter because it is doing too much. According to our model, disappointment with the government will come especially from the more skeptical ones, who are expected to resist policy implementation. This will feed back into higher social distrust, especially among those who are more worried. The stricter the rules are perceived to be, or the larger the costs of complying with them, the more these effects will play out.

5 Concluding Remarks

The ideas developed in this paper can be extended in different directions providing a stepping-stone for a deeper understanding of why distrust is increasing in advanced societies, how it translates into the political behavior of individuals, and what strategies can be implemented to rebuild trust in society.

Voters, for example, can be more easily mobilized towards protest voting when they distrust institutions. Political leaders can manipulate their opinions to generate social and political distrust, demonize their counterparts, and polarize political confrontation. In a society where minority and majority do not trust each others, the latter may have less hesitation in approving laws that are particularly burdensome for the minority. The relationship between distrust, partisan identification, and polarization of political preferences is important and requires further investigation.

What can the government do to limit the impact of regulation on citizens' distrust? Answering this question requires a formal normative analysis, which we leave to future work. It could be done by introducing a cost of distrust into the government's objective function. Arguably, taking into account the cost of distrust would lead the government to abstain from regulating some types of externalities, especially when disagreement about

them is large. Moreover, if citizens' views were skewed to one extreme or the other, a government willing to compromise between its view of the "best policy" and the effect on distrust would lean toward them. This could result in a loss of material welfare for society as a whole. Of course, a huge loss of social trust would occur any time the government introduced new rules but failed to enforce them. This would have the effect of increasing the perception of uncivic behavior in society. Conversely, overly restrictive regulation would result in a large loss of institutional trust, as many would find the rules useless and end up blaming the government in the first place.

Our framework could be extended, with different implications, by imagining that individuals, in addition to having different opinions, may also have heterogeneous disposition to dialogue and to the exchange of ideas. It would be realistic to hypothesize that exchanging ideas leads to compromise in a world where individuals trust political institutions and others, while it leads to confrontation and conflict in a world where people distrust institutions and others. Trust would therefore be a fundamental ingredient of democratic dialogue and would be fed endogenously precisely by democratic dialogue. Conversely, contrast and juxtaposition of ideas breeds distrust, as we have described in our model, and this mechanism would endogenously lead to further distrust.

Information plays a key role. The abundance of non-scientific and distorted sources feeds the incidence of cognitive biases and the risk of polarization of preferences, which in turn yields distrust. Our model does not offer a strategy to rebuild trust. However, it does suggest that reducing non-dialectical opposition of political opinions is a crucial step in preventing trust erosion in advanced democracies.

References

Aghion, P., Algan, Y., Cahuc, P., and Shleifer, A. (2010). Regulation and distrust. *The Quarterly journal of Economics*, 125(3):1015–1049.

Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Alesina, A., Miano, A., and Stantcheva, S. (2020). The polarization of reality. *American Economic Review Papers and Proceedings*, 110:324–328.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.

Alesina, A., Tabellini, G., and Trebbi, F. (2017). Is europe an optimal political area? *Brookings Papers on Economic Activity, SPRING*, pages 169–213.

Algan, Y. and Cahuc, P. (2009). Civic virtue and labor market institutions. *American Economic Journal: Macroeconomics*, 1(1):111–45.

Algan, Y., Cohen, D., Davoine, E., Foucault, M., and Stantcheva, S. (2021). Trust in scientists in times of pandemic: Panel evidence from 12 countries. *Proceedings of the National Academy of Sciences*, 118(40).

Canen, N., Kendall, C., and Trebbi, F. (2020). Unbundling polarization. *Econometrica*, 88(3):1197–1233.

Clarke, D., Romano, J. P., and Wolf, M. (2020). The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, 20(4):812–843. Publisher: SAGE Publications.

Coleman, J. S. (1994). *Foundations of social theory*. Harvard university press.

de Quidt, J., Johannes, H., and Christopher, R. (2018). Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11):64.

Dechezleprêtre, A., Fabre, A., Kruse, T., Planterose, B., Chico, A. S., and Stantcheva, S. (2022). Fighting climate change: International attitudes toward climate policies. Technical report, National Bureau of Economic Research.

Delhey, J. and Newton, K. (2005). Predicting cross-national levels of social trust: global pattern or nordic exceptionalism? *European sociological review*, 21(4):311–327.

Delhey, J., Newton, K., and Welzel, C. (2011). How general is trust in "most people"? solving the radius of trust problem. *American sociological review*, 76(5):786–807.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1):143–181.

Durante, R., Guiso, L., and Gulino, G. (2021). Asocial capital: Civic culture and social distancing during covid-19. *Journal of Public Economics*, 194:104342.

Gecewicz, C. and Rainie, L. (2019). Why americans don't fully trust many who hold positions of power and responsibility. *Pew Research Center Report*.

Gentzkow, M. (2016). Polarization in 2016. *Toulouse Network for Information Technology Whitepaper*, pages 1–23.

Gitmez, A., Sonin, K., and Wright, A. L. (2020). Political economy of crisis response.

University of Chicago, Becker Friedman Institute for Economics Working Paper, (2020-68).

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2):23–48.

Guiso, L., Sapienza, P., and Zingales, L. (2011). Civic capital as the missing link. *Handbook of social economics*, 1:417–480.

Hardin, R. (2002). *Trust and trustworthiness*. Russell Sage Foundation.

Hardin, R. (2006). *Trust*, volume 10. Polity.

Knack, S. and Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3):207–227.

Levi, M. and Stoker, L. (2000). Political trust and trustworthiness. *Annual review of political science*, 3(1):475–507.

Martinangeli, A. F. M., Povitkina, M., Jagers, S., and Rothstein, B. (2024). Institutional Quality Causes Generalized Trust: Experimental Evidence on Trusting under the Shadow of Doubt. *American Journal of Political Science*, 68(3):972–987. https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12780.

Newton, K., Stolle, D., and Zmerli, S. (2018). Social and political trust. *The Oxford handbook of social and political trust*, 37:961–976.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997. *American Political Science Review*, 92(1):1–22.

Passarelli, F. and Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3):903–946.

Pinotti, P. (2012). Trust, regulation and market failures. *Review of Economics and Statistics*, 94(3):650–658.

Putnam, R. (1993). The prosperous community: Social capital and public life. *The American Prospect*, 13(4).

Putnam, R. D. et al. (2000). *Bowling alone: The collapse and revival of American community*. Simon and schuster.

Romano, J. P. and Wolf, M. (2005). Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Sta-*

tistical Association, 100(469):94–108. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/016214504000000539.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404.

Stantcheva, S. (2021). Understanding tax policy: How do people reason? *Quarterly Journal of Economics*, 136(40):2309–2369.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15:205–234.

Stevenson, B. and Wolfers, J. (2011). Trust in public institutions over the business cycle. *American Economic Review*, 101(3):281–87.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of europe. *Journal of the European Economic association*, 8(4):677–716.

Yamagishi, T. (2011). *Trust: The evolutionary game of mind and society*. Springer.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Appendices

For Online Publication

A Proofs

Proof of Lemma 1. 1. For all Group 1's members, the equilibrium behavior satisfies equation (5). Because a stricter rule (holding fixed monitoring) and stricter monitoring (holding fixed the rule) lead to higher average behavior in society, the equilibrium behavior of any individual in Group 1 is decreasing in ρ and m, by strategic substitutability. The lowest unaffected type in this group is the one choosing to behave as ρ even if she is not fined. Thus, this type, $x(\rho, m)$, is defined by

$$\varepsilon'_{x(.)}(\rho + B^*_{-x(.)}(\rho, m)) - c'(\rho) = 0 \tag{17}$$

Implicitly differentiating equation (17) with respect to ρ and using the envelope theorem yields

$$\frac{\partial x}{\partial \rho} = -\frac{\varepsilon''_x(\rho + B^*_{-x}(.)) \cdot (1 + \frac{\partial B^*_{-x}(.)}{\partial \rho}) - c''(\rho)}{\frac{\partial \varepsilon'_i(\rho + B^*_{-i}(.))}{\partial i}} > 0$$

where $\frac{\partial B^*_{-i}(.)}{\partial \rho}$ in the first term at the numerator accounts for those individuals whose equilibrium behavior is in a corner.

Similarly, implicitly differentiating equation (17) wrt m yields

$$\frac{\partial x(.)}{\partial m} = -\frac{\varepsilon''_x(\rho + B^*_{-x}(.)) \cdot \frac{\partial B^*_{-x}(.)}{\partial m}}{\frac{\partial \varepsilon'_i(\rho + B^*_{-i}(.))}{\partial i}} > 0.$$

Therefore, $x(\rho, m)$ positively depends on ρ and/or m. In other words, the size of Group 1 is smaller when the rule is more stringent and/or monitoring is more intense.

2. For individuals in Group 2, (6) is a strict inequality. Their optimal behavior ρ is in a corner. By definition their behavior is unaffected by m, and it is increasing in ρ.

3. For individuals in Group 3, expression (6) holds with equality. Those choosing not to comply reduce their behavior because of strategic substitutability. This can be shown by implicitly differentiating equation (6) using the envelope theorem and confirming that the implicit derivative of behavior with respect to the rule is negative, as follows

$$\frac{\partial b_i^*}{\partial \rho} = -\frac{\varepsilon_i''(b_i^*(.) + B_{-i}^*(.)) \cdot \frac{\partial B_{-i}^*(.)}{\partial \rho} - \overline{k} \cdot \frac{\partial^2 p(..)}{\partial b_i^* \partial \rho}}{\varepsilon_i''(b_i^*(.) + B_{-i}^*(.)) \cdot \frac{\partial B_{-i}^*(.)}{\partial b_i^*} - c''(b_i^*(.)) - \overline{k} \cdot \frac{\partial^2 p(..)}{\partial b_i^{*2}}} < 0.$$

where the inequality follows as a sufficiency result from the fact that individuals in the second group increase their behavior, so that $\frac{\partial B_{-i}^*(.)}{\partial \rho} > 0$, and the assumption that the marginal impact of behavior on detection probability $p(..)$ does not depend on the strictness of the rule, and thus $\frac{\partial^2 p(..)}{\partial b_i^* \partial \rho} = 0$.

Following the same steps,

$$\frac{\partial b_i^*}{\partial m} = -\frac{\varepsilon_i''(b_i^*(.) + B_{-i}^*(.)) \cdot \frac{\partial B_{-i}^*(.)}{\partial m} - \overline{k} \cdot \frac{\partial^2 p(..)}{\partial b_i^* \partial m}}{\varepsilon_i''(b_i^*(.) + B_{-i}^*(.)) - c''(b_i^*(.)) - \overline{k} \cdot \frac{\partial^2 p(..)}{\partial b_i^{*2}}} > 0$$

Here the result follows from our assumption that the marginal impact of behavior on detection probability is large enough and gets stronger if m is higher and thus, $\frac{\partial^2 p(..)}{\partial b_i^* \partial m} < 0$. This assumption entails that the sign of the numerator is determined by the second term. In words, the incentive to increase behavior due to the higher marginal impact on detection probability (i.e., the second term) is stronger than the incentive to lower behavior by strategic substitutability (the first term). This implies that individuals in the third group will increase their behavior, and hence the individuals in the first groups can reduce theirs (by strategic substitutability). Therefore, the government can use monitoring to induce more mitigating behavior on average in the population, for any value of ρ.

The highest type in Group 3 is $z(\rho, m)$, defined by the following equation

$$\varepsilon'_{z(.)}(\rho + B_{-z(.)}^*(\rho, m)) - c'(\rho^*) - \overline{k} \cdot \left.\frac{\partial p(b_i^*(.), \rho, m)}{\partial b_i^*}\right|_{b_i^* = \rho} = 0 \tag{18}$$

Implicitly differentiating (18) with respect to ρ and using the envelope theorem yields

$$\frac{\partial z}{\partial \rho} = -\frac{\varepsilon_z''(\rho + B_{-z}^*(.)) \cdot (1 + \frac{\partial B_{-z}^*(.)}{\partial \rho}) - c''(\rho) - \overline{k} \cdot \frac{\partial^2 p(..)}{\partial b_i^* \partial \rho}}{\frac{\partial \varepsilon_i'(\rho + B_{-i}^*(.))}{\partial i}} > 0$$

where the inequality follows as a sufficiency result from our assumption that ρ has no impact on the marginal effect of behavior on detection probability, and thus $\frac{\partial^2 p(..)}{\partial b_i^* \partial \rho} = 0$.

■

Proof of Proposition 1. We derive the proof in two steps in what follows.

i. For notational simplicity, let $\varepsilon_i(\rho) \equiv \varepsilon_i(b_i^*(\rho, 1) + B_{-i}^*(\rho, 1))$, and $c(\rho) \equiv c(b_i^*(\rho, 1))$, for all i, ρ, and $m = 1$.

Take any symmetric couple of l and h with $\rho_l^* < \rho_g^* = \rho_G^* < \rho_h^*$ and $|g - l| = |g - h|$. By (15),

$$D_l^I(\rho_G^*) = -\int_{\rho_l^*}^{\rho_G^*} \varepsilon_l'(\rho) - c'(\rho) d\rho$$

and

$$D_h^I(\rho_G^*) = \int_{\rho_G^*}^{\rho_h^*} \varepsilon_h'(\rho) - c'(\rho) d\rho$$

By the two definite integrals above, proving that $D_l^I(\rho_G^*) > D_h^I(\rho_G^*)$ implies proving that

$$[\varepsilon_h(\rho_h^*) - \varepsilon_h(\rho_G^*)] - [c(\rho_h^*) - c(\rho_G^*)] < [c(\rho_G^*) - c(\rho_l^*)] - [\varepsilon_l(\rho_G^*) - \varepsilon_l(\rho_l^*)] \qquad (19)$$

The first square brackets on the LHS represents the *loss of benefits* from externality containment borne by the high-concern type when rule ρ_G^* is implemented instead of the policy entitlement, ρ_h^*. The second one is the gain resulting from *cost savings*. This gain partly offsets the loss, thereby contributing to a reduction in distrust for type h, $D_h^I(\rho_G^*)$. Note that the first term is always larger than the second one. However, if $\varepsilon_h(\rho)$ is a highly concave function the first term (i.e., loss of benefits) is a relatively small number; additionally if $c(\rho)$ is a highly convex function then the second term (i.e., the partially offsetting gain) is a relatively large number. Overall, concavity of the benefit function and/or convexity of the cost function yield a relatively low value of the LHS, thus resulting in a relatively small level

of distrust for the high-concern type h. The first square brackets on the RHS represents the loss in terms of higher costs borne by the low-concern type when rule ρ_G^* is enacted instead of her counterfactual entitled rule, ρ_l^*. The second brackets on the RHS represents the gain in terms of externality mitigation. The latter partly offsets the loss, thereby mitigating distrust, $D_h^I(\rho_G^*)$. However, if $\varepsilon_h(\rho)$ is a highly concave function this mitigating effect is small, whereas if $c(\rho)$ is a highly convex function then the loss is significant. Overall, concavity of the benefit function and/or convexity of the cost function yield a relatively high value of the RHS, and result in a relatively large level of distrust for the low-concern type l. This can be more clearly understood by manipulating inequality (19), as follows:

$$[\varepsilon_l(\rho_G^*) - \varepsilon_l(\rho_l^*)] + [\varepsilon_h(\rho_h^*) - \varepsilon_h(\rho_G^*)] < c(\rho_h^*) - c(\rho_l^*)$$

Here the first term on the LHS is a small number because $\varepsilon_l(\cdot)$ is a function assigning low values to its arguments. The second term is a small number too if the benefit function $\varepsilon_h(\cdot)$ is quite concave. The RHS is a large number if $c(\cdot)$ is a quite convex function. Overall the inequality holds if the benefit functions are quite concave and the cost function is sufficiently convex. Therefore, enough convexity of the cost function and enough concavity of the benefit functions of all individuals will lead to $D_l^I(\rho_G^*) > D_h^I(\rho_G^*)$ for any couple of symmetric individuals, l and h, and for any ρ_G^*. The reverse is also true.

ii. a) Take an unaffected individual i in Group 1, as defined by Lemma 1: $b_i^*(\rho_G^*, m_G^*) > \rho_G^*$. By (16) and (2),

$$\begin{aligned} D_i^S &= \left[\varepsilon_i\left((b_i^*(\rho_G^*,1) + \int_0^{z(.)} b_j^*(\rho_G^*,1) + \int_{z(.)}^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*,1)dj\right)\right. \\ &- \left.\varepsilon_i\left((b_i^*(\rho_G^*,m_G^*) + \int_0^{z(.)} b_j^*(\rho_G^*,m_G^*) + \int_{z(.)}^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*,m_G^*)dj\right)\right] \\ &+ \left[c(b_i^*(\rho_G^*,m_G^*)) - c(b_i^*(\rho_G^*,1))\right]. \end{aligned}$$

The first squared brackets represents the loss of benefits due to misbehavior of non-compliant people below $z(.)$. It is positive and increasing in i's level of concern. The second squared brackets is the cost increase due to the fact that, because of strategic

substitutability, when some people do not comply individual i has to increase her behavior in equilibrium. Also this second term is increasing in i's level of concern because, due to concavity of $\varepsilon_i(.)$, a higher type has to increase his behavior by a larger amount given the misbehavior of non-compliant people.

b) Now consider an affected individual i in Group 2, as defined by Lemma 1: $b_i^*(\rho_G^*, m_G^*) = \rho_G^*$. In this case,

$$\begin{aligned} D_i^S &= \left[\varepsilon_i\left(\rho_G^* + \int_0^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*, 1)dj\right)\right] \\ &- \left[\varepsilon_i\left(\rho_G^* + \int_0^{z(.)} b_j^*(\rho_G^*, m_G^*) + \int_{z(.)}^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*, m_G^*)dj\right)\right]. \end{aligned}$$

Social distrust is only the loss of benefits due to misbehavior of all non-compliant people in $[0, z(.))$. The higher i, the bigger $\varepsilon_i(.)$, the larger that loss of benefits.

c) Finally consider a non-compliant individual $i \in [0, z(.))$ in Group 3, with $b_i^*(\rho_G^*, m_G^*) < \rho_G^*$ (see Lemma 1). She can only complain that some lower types $j \in [0, i)$ are less compliant than she is. This causes the following loss of benefits:

$$\begin{aligned} D_i^S &= \left[\varepsilon_i\left(b_i^*(\rho_G^*, m_G^*) + \int_0^i b_i^*(\rho_G^*, m_G^*)dj + \int_i^{z(.)} b_j^*(\rho_G^*, m_G^*)dj + \int_{z(.)}^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*, m_G^*)dj\right)\right] \\ &- \left[\varepsilon_i\left(b_i^*(\rho_G^*, m_G^*) + \int_0^{z(.)} b_j^*(\rho_G^*, m_G^*)dj + \int_{z(.)}^{x(.)} \rho_G^* dj + \int_{x(.)}^1 b_j^*(\rho_G^*, m_G^*)dj\right)\right] \\ &+ \left[c(b_i^*(\rho_G^*, m_G^*)) - c(b_i^*(\rho_G^*, m_G^*)|_{b_j^* \geq b_i^*})\right]. \end{aligned}$$

In this case, total counterfactual behavior of lower types $j \in [0, i)$ is $\int_0^i b_i^*(\rho_G^*, m_G^*)dj$ (i.e., in the counterfactual, all lower types should behave as i does), while their total actual behavior is $\int_0^i b_j^*(\rho_G^*, m_G^*)dj$. The higher i, the larger the difference between the counterfactual and the actual total behavior of lower types. Moreover, the higher i, the larger the loss of benefits associated with that difference. Both factors lead to bigger social distrust. In addition, the last squared brackets, analogous to case a), is the cost increase due to the fact that in the counterfactual $b_i^*(\rho_G^*, m_G^*)|_{b_j^* \geq b_i^*}$ is lower than i's actual behavior $b_i^*(\rho_G^*, m_G^*)$, because of strategic substitutability. Also this second term is increasing in i's level of concern.

Then, a) plus b) plus c) prove statement ii.

■

Proof of Proposition 2. We sketch the first part of the proof in words.

i. We capture higher disagreement with higher mean-preserving dispersion of $\epsilon_i'(.)$, where the mean is $\epsilon_a'(.)$. With more disagreement low-concern individuals are less concerned and want lower policies, while high-concern individuals are more concerned and want stricter policies. For any government's policy $\rho_G^* = \rho_a^*$, the distance between individuals' preferred rules, ρ_i^* and ρ_G^* will (weakly) increase for all i, leading to bigger institutional distrust.

   As for social distrust, holding ρ_G^* constant, some low-concern individuals are less worried and have stronger incentives to lower their behavior, while high-concern individuals are more worried and have stronger incentives to increase their behavior. There will be more non-compliant people which in turn will be blamed more by those who are more concerned. This will increase total interpersonal distrust in society.

ii. With no big loss of generality, assume concern increases for all individuals by a proportion $\gamma_i > 0$ so that i's level of concern is $\varepsilon_i'(b_i + B_{-i})(1+\gamma_i)$, for any i and any value of $(b_i + B_{-i})$. We can then rewrite equation (10) at the ideal rule as $\varepsilon_i'(\rho_i^* + B_{-i}^*(\rho_i^*, 1))(1+\gamma_i) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right) - c'(\rho_i^*) = 0$. Implicitly differentiating it yields,

$$\frac{\partial \rho_i^*}{\partial \gamma_i} = -\frac{\varepsilon_i'(..) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right)}{\varepsilon_i''(..)(1+\gamma_i) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right)^2 + \varepsilon_i'(.)(1+\gamma_i) \cdot \frac{\partial^2 B_{-i}^*(\rho,1)}{\partial \rho^2} - c''(.)} > 0,$$

   so that the more individuals are concerned the more restrictive they want the policy.

   Second-order implicit differentiation yields

$$\frac{\partial^2 \rho_i^*}{\partial \gamma_i^2} = \frac{\left[\varepsilon_i'(..) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right)\right] \cdot \left[\varepsilon_i''(.) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right)^2 + \varepsilon_i'(.) \cdot \frac{\partial^2 B_{-i}^*(\rho,1)}{\partial \rho^2}\right]}{\left(\varepsilon_i''(..)(1+\gamma_i) \cdot \left(1 + \frac{\partial B_{-i}^*(\rho,1)}{\partial \rho}\right)^2 + \varepsilon_i'(.)(1+\gamma_i) \cdot \frac{\partial^2 B_{-i}^*(\rho,1)}{\partial \rho^2} - c''(.)\right)^2}.$$

   The denominator is positive. The first squared bracketed term in the numerator is positive. The second one is positive if $\varepsilon_i'(\rho_i^* + B_{-i}^*(\rho_i^*, 1)) > 0$ is sufficiently large

for all i. In this case $\frac{\partial^2 \rho_i^*}{\partial \gamma_i^2} > 0$ and the positive effect of an increase in concern on entitled policy is larger for high-concern individuals (who want a higher ρ_i^*) compared to low-concern individuals. In other words, entitled rules increase more than proportionally with types. Because ρ_G^* equals the most preferred rule of the average-concern individual, it will increase as well. However, since ρ_i^* increases more than proportionally with i, the distance between individuals' entitled policy and the government's policy increases for all individuals. This implies that the loss of not implementing the entitled policy is higher for all individuals. Thus total social distrust increases too. Note that this proof is based on a sufficient condition, implying that institutional distrust increases for all individuals. The statement would hold even if $\frac{\partial^2 \rho_i^*}{\partial \gamma_i^2} > 0$ holds only for a sufficiently large subset of individuals, rather than the entire population. Also the loss due to bad behavior of others increases because now virtuous behavior is valued more by all individuals. Thus total social trust increases too.

Both points prove the proposition. ■

Proof of Proposition 3. Following the proposition, the proof is split in two parts:

i. We do not provide a formal proof, but we sketch it and provide the intuition. By equation (15), institutional distrust increases in the distance between i's entitled rule and government's rule. We then consider three groups of individuals. A first group is made up of all individuals whose entitled policy is lower than the government's policy: $\rho_G^* - \rho_i^* > 0$. The increase in ρ_G^* results in a higher distance and stronger distrust. A second group consists of sufficiently high-concerned individuals, so that the increase in ρ_G^* results in a lower distance and lower distrust. Finally, there is a third group of "intermediate" types. For these individuals $\rho_i^* - \rho_G^*$ turns from positive to negative. They turn from distrusting institutions because the policy was too low to distrusting it because the policy is now too high. Some of those individuals will distrust institutions more compared to the initial situation, while some others will distrust institutions less. If the increase in ρ_G^* is sufficiently large, the number of those in the first group and in the third group distrusting the government more is sufficiently large so that total institutional distrust increases.

ii. By Lemma 1 recall that equation (6) taken as equality pins down the lowest complying individual, $z(\rho_G^*, m_G^*)$. By Lemma 1, $\frac{\partial z}{\partial \rho_G^*} > 0$, which implies that a higher ρ_G^* leads less individuals to fully comply. Moreover, by Lemma 1, $\frac{\partial b_i^*}{\partial \rho_G^*}\Big|_{b_i^* < \rho_G^*} < 0$. Those choosing not to comply reduce their behavior because of strategic substitutability. Both points prove the proposition.

■

Proof of Proposition 4. Following the proposition, the proof is split in two parts:

i. We do not provide a formal proof, but we sketch it. By equation (10) all individuals' entitled policies are lower if marginal costs are higher. As in Proposition 3.i all low-concerned individuals whose entitled policy is lower than ρ_G^* distrust the government more because the distance between their entitled policies and ρ_G^* is bigger. All high-concerned individuals whose entitled policy is higher than ρ_G^* distrust the government less because the distance between their entitled policies and ρ_G^* shrinks. The function mapping types into institutional distrust shifts up for low-concerned types and down for high-concerned types. However, by Proposition 1.i a high-concerned type experiences a lower feeling of institutional distrust compared to a low-concerned individual who is symmetric with respect to the type preferring ρ_G^*. Take any high-concerned type h who is symmetric with respect to type l, so that $\rho_h - \rho_G^* = \rho_G^* - \rho_l$, by Proposition 1.i we know that $D_h^I < D_l^I$. As costs increase, all individuals' preferred rules shift downwards. Let ρ_h' be the new rule preferred by individual h. The individual preferring a symmetric rule is now $l1$, with $\rho_h' - \rho_G^* = \rho_G^* - \rho_{l1} < \rho_G^* - \rho_l$. By Proposition 1.i we have that $D_h'^I < D_{l1}'^I$ and we also have that $l < l1$, so that $D_h'^I < D_{l1}'^I < D_l'^I$. This holds for any h and any l implying that the total increase of distrust is overall higher for low-concerned individuals and proves the proposition.

ii. Using Lemma 1 and following the same steps as in the proof of Proposition 3, if marginal costs increase, all non-compliant individuals will decrease their behavior and some compliant individuals will choose not to comply any more (i.e. the lowest complying type, $z(.)$, increases) . High-types with $b_i^* \geq \rho_G^*$ will distrust others more. All non-compliant individuals with $b_i^* < \rho_G^*$ will distrust individuals j with $b_i^* > b_j^*$ because their behavior is lower.

Both points prove the proposition. ■

B The Survey

We hired the survey company 'Respondi' to simultaneously distribute the link to our survey in Germany, Italy, the Netherlands and Spain in the first two weeks of June 2020.[38] We collected data from a random sample of adults (above 17 and under 70 years of age) exceeding 2000 individuals per country (see Appendix C.7 for details).[39] We aimed at representativeness of the samples by age, geographic area of residence and gender, and targeted a distribution of disposable equivalent household income as close as possible to the one available in Eurostat.[40] The English questionnaire (link in Appendix D) was translated by the native-speaking authors, except for the Spanish version which was instead translated by Respondi, and was administered in the local language (links to the surveys in Appendix D).

At the very beginning of the survey we elicit the respondents' gender, age, marital status, household size (number of adults and number of children), household monthly disposable income.

At the very end we further collect information about highest educational attainment, media information sources, employment status, immigration background, political beliefs and turnout at the last election.

B.1 Incentivised Willingness to Support European Union

To better capture the epidemic's impact on the respondents' attitudes towards the European Union, we include an incentivised measure of their willingness to engage in an action explicitly framed as supportive of the European integration project. The respondents read that "*For educational purposes, we are considering informing students about the importance of the European Union using real texts. We selected a speech promoting European integration which was given in front of the European Parliament.*"[41]

[38] https://www.respondi.com/EN/

[39] We are a priori able to detect a minimum effect MDE=0.12 on standardised outcome measures at $\alpha = 0.05$ and power $\pi = 0.8$ in within-country analyses.

[40] EU-SILC: https://ec.europa.eu/eurostat/web/main/home

[41] For an English transcription of the speech: https://www.elysee.fr/emmanuel-macron/2018/04/17/speech-by-emmanuel-macron-president-of-the-republic-at-european-parliament.en

We then ask the respondents whether they would be willing to read a five-minute long transcription of the speech, delivered in 2017, and to give us their opinion about the suitability of the text for the purpose it was selected for. We thus provide a clear incentive to decline to respondents who are *not willing* to spend five minutes (it took on average approximately 20 minutes to complete the survey *without* reading the text) to read a pro-European Union text (see also DellaVigna et al. (2016)). We explicitly fixed the amount of time needed to read the speech in order to fix beliefs about the length of the task and the amount of time and effort needed to complete it. Further, the explicit reference to the educational usage of the text (in a Public Economics undergraduate course at the University of Stirling taught by one of the authors) serves the purpose of providing the respondents with a sense of consequentiality of the action. The identity of the speaker (Emmanuel Macron) and the context in which the speech was given was not disclosed to the respondents at the time of their choice. We informed the respondents that their agreement or lack thereof will not affect their payment. In case of agreement, the respondents were told that they would read and review the text only at the end of the survey.[42]

We interpret the respondents' choice of (not) reading the text as (un)willingness to support the European integration, and *not* the rating provided: It might be that a respondent with positive attitudes towards the European integration might legitimately find the text unsuitable for the purpose and assign it a low rating. A non-trivial choice was whether to explicitly frame the action as supportive of European integration or whether to maintain a neutral wording. In the latter case, however, the interpretation of the agreement to read the text would have not been straightforward. As argued above, framing it as pro-integration allows for agreement to read to be interpretable as supportive of the European integration, irrespective of the rating assigned to the speech. Neutral wording would have not allowed this interpretation. It can be argued that our incentivised measure of support for the European integration could have in such case been the rating distribution. Notice however that, in addition to the arguments above that a low rating does not straightforwardly translate into low support for the European Union, respondents choosing not to read the text would have been dropped out of the analysis. Moreover, the incentivising intent would have been lost (it is costly to spend five more

[42]For the experimental challenges posed by this question and how they are addressed, see Appendix B.3.

minutes to read but it is costless to assign the rating). We cannot completely exclude that the respondents might accept to read and then assign ratings without reading. The incidence of such behaviors is however likely to be orthogonal to our design and smaller than with neutral wording. Our choice moreover allows us to perform analyses allowing us to gauge the validity of the responses collected and of our incentivised measure.

B.2 The Text Agreement Question: Behavioral Analyses

The analyses here presented follow the analytical framework outlined in Section 3 and confirm the validity of our behavioral measure of support for the EU. In order to obtain a more realistic picture, we trim the data by excluding from the following analyses the upper tail of the distribution of time spent reading text: the top 1%. These are respondents who spent half an hour or more on the text screen. The respondents spent on average 213 seconds (slightly short of 4 minutes) on the text screen, with little variation across conditions (Baseline: 204.56, SD=248.75; : 218.03, SD=258.04; Health: 217.57, SD=264.42; Economic: 213.72, SD=242.4; Conflict: 222.72, SD=266.34).

We use OLS analyses to look for differences in the amount of time spent reading the text across the conditions and the baseline. Differences in time spent on the text are mostly not significant at conventional levels, and where significant they are small in magnitude (vs Baseline: $\beta = 13.32$, SE=5.858; 'Health' vs Baseline: $\beta = 10.84$, SE=8.115; 'Health & Economic' vs Baseline: $\beta = 9.23$, SE=7.732; 'Health & Conflict' vs Baseline: $\beta = 19.58$, SE=9.268).

We moreover compare the distributions of the text ratings across conditions using two-sided Kolmogorov-Smirnov tests. The distributions are extremely similar in all cases: null hypothesis of equality of the populations cannot be rejected in three out of four comparisons (vs Baseline, p-value=0.118; Health vs Baseline, p-value=0.060; Economic vs Baseline, p-value=0.535; Conflict vs Baseline, p-value=0.867).

Finally, we investigate whether a relationship exists between the rating assigned to the text and the time spent reading it among those who chose to do so. An OLS regression reveals that respondents who assigned a greater rating also spent a significantly larger amount of time in seconds reading the text ($\beta = 0.000814$, SE=0.000151, p-value<

0.01).[43] Notice however that though precisely estimated, the coefficient is small: an additional 30 seconds increases the score by 0.02 points.

B.3 Challenges of the Experimental Design

Fatigue Fatigue might influence the propensity to review our text on European integration. Recall that half of the respondents will receive the question on whether they wish to read a lengthy text (explicitly fixed at 5 minutes of time) about the European integration relatively early in the survey, while half will receive it relatively late due to the randomization into and Baseline condition. Among the latter, greater fatigue is expected to decrease the likelihood of agreement. For this reason, we treat fatigue as a confound deserving high priority.

The text agreement question is therefore randomly placed at the beginning or at the end of the outcomes block: its placement varies therefore between early on, somewhat in the middle and towards the end of the entire survey. In case of agreement, the respondents will read the text and provide their opinion at the end of the questionnaire, shielding the following parts of the survey from additional fatigue.

Experimenter demand effects Participants to surveys or experiments might infer the researchers' underlying objectives from the questions asked and/or from the experiment's architecture, and act to comply with what they believe are the experimenter's objectives Zizzo (2010). It is unclear in which direction our respondents would infer our objectives to lie. First of all, each respondent is blind to all conditions and sub-conditions besides the one they are taking part in and hence cannot use the experimental design to make such inference. Second, it remains unclear why a respondent in the COVIDFIRST condition would systematically infer they should favour a direction over another. Both uncovering stronger or weaker trust, stronger or weaker support for governments and state interventions are, from the respondents' viewpoint equally compelling and likely objectives of this study.

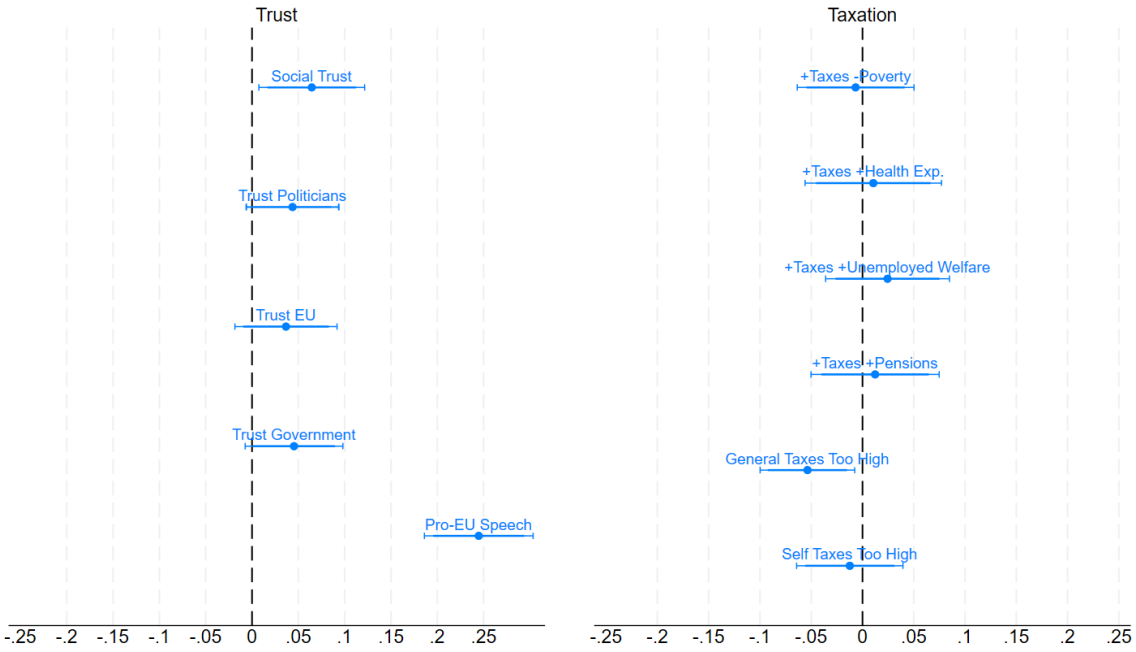
We can evaluate the likelihood of demand effects polluting our questionnaire by ex-

[43] Controls include gender, age groups, employment status, education, immigrant status, family status and number of family members, equivalised household income (coded into five quantiles), a dummy to define the position of the Macron Speech question (see Section B.3 for more details) and regional fixed effects.

ploiting the randomization of the position of the text agreement question. The explicit pro-EU sentiment in that question leads to a strengthened pro-institutional demand effect affecting subsequent questions. This fact allows us to evaluate how 'explicitly pushing' respondents towards inferring our objectives along the lines of what suggested by de Quidt et al. (2018). Comparing the responses of those exposed to the text agreement at the beginning and at the end of the outcomes block allow us to establish whether the survey is susceptible to any demand effect.

Figure B1 illustrates the coefficients associated to a dummy variable indicating whether the respondents were asked whether they wished to read the text about the European Union integration *before* or *after* they had answered our target outcome questions. First, and unsurprisingly, respondents receiving the text agreement question at the beginning of the survey, are more likely to read it. Second, we find a statistically significant increase in social trust and, almost significant coefficients for trust in politicians and the national government. Similarly, respondents receiving the text agreement question at the beginning are less likely to reply that taxes are too high. Overall, we cannot exclude an experimenter demand effects on some of the outcomes. We tackle this potential issue by: i) in all regressions, we control for the position of the text agreement question in the survey, as the demand effect is determined by the placement early or in the middle of the survey; ii) in a robustness test, we control for the interaction between the dummy *Covidfirst* and the dummy *Beginning*, which takes values 1 for an early placement of the text agreement question. The findings are displayed in Tables B1, B2 and B3. The coefficients of interest (*Covidfirst * Health Concerns*, *Covidfirst * Compliance*, and *Covidfirst * Pandemic Behaviors*) are barely affected and remain always statistically significant and similar in size. The interaction *Covidfirst * Beginning* seems to matter only for the Pro-EU speech.

Figure B1: Effect of the position of the text agreement question



The figure displays the impact of the position of the text agreement question: at the very beginning or at the very end of the socio-political attitudes block. The figure also displays 95% (delimited by vertical bars) and 90% (bold line) confidence intervals.

Table B1: Priming Covid-19 and Demand Effects - Health Concerns

	(1) Trust Others	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Pro EU Speech	(5) Trust EU
Covidfirst	-0.118***	-0.007	-0.036	-0.071*	-0.135***
	(0.040)	(0.032)	(0.030)	(0.038)	(0.033)
Health Concerns	-0.024	0.067***	0.068***	0.031**	0.109***
	(0.022)	(0.022)	(0.022)	(0.015)	(0.023)
Covidfirst*Health Concerns	-0.041*	0.058**	0.060***	0.038**	0.007
	(0.024)	(0.025)	(0.023)	(0.019)	(0.029)
Beginning = 1	0.068*	0.027	0.041	0.392***	0.005
	(0.041)	(0.031)	(0.033)	(0.037)	(0.036)
Covidfirst*Beginning	-0.002	0.028	0.010	-0.147***	0.038
	(0.050)	(0.043)	(0.042)	(0.052)	(0.045)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.079	0.156	0.214	0.088	0.058
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Health Concerns': self-perceived Covid-19 health risks from 1 (very low) to 10 (very high). 'Beginning': text agreement question at the beginning of the survey. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table B2: Priming Covid-19 and Demand Effects - Compliance

	(1) Trust Others	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Pro EU Speech	(5) Trust EU
Covidfirst	-0.119***	-0.004	-0.033	-0.070*	-0.130***
	(0.040)	(0.031)	(0.030)	(0.038)	(0.033)
Compliance	0.015	0.098***	0.084***	-0.003	0.120***
	(0.021)	(0.022)	(0.023)	(0.020)	(0.024)
Covidfirst*Compliance	-0.064***	0.067***	0.050**	0.043**	0.021
	(0.024)	(0.022)	(0.020)	(0.022)	(0.027)
Beginning = 1	0.068*	0.030	0.044	0.392***	0.010
	(0.041)	(0.031)	(0.033)	(0.037)	(0.036)
Covidfirst*Beginning	-0.002	0.030	0.010	-0.147***	0.037
	(0.051)	(0.043)	(0.042)	(0.052)	(0.045)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.078	0.162	0.214	0.086	0.061
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Compliance': level of compliance with Covid-19 regulations from 1 (very low) to 10 (very high). 'Beginning': text agreement question at the beginning of the survey. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table B3: Priming Covid-19 and Demand Effects - Pandemic Behaviors

	(1) Trust Others	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Pro EU Speech	(5) Trust EU
Covidfirst	-0.065	0.024	-0.012	-0.147**	-0.063
	(0.051)	(0.052)	(0.047)	(0.072)	(0.054)
Civic Expectations	0.033	0.192***	0.140***	-0.036	0.151***
	(0.029)	(0.031)	(0.033)	(0.041)	(0.035)
Covidfirst*Civic Expectations	-0.112***	0.050	0.029	0.052	0.022
	(0.039)	(0.040)	(0.040)	(0.048)	(0.044)
Beginning = 1	0.063	-0.012	-0.005	0.269***	-0.010
	(0.064)	(0.064)	(0.065)	(0.077)	(0.069)
Covidfirst*Beginning	0.000	0.026	0.033	-0.004	-0.011
	(0.085)	(0.084)	(0.077)	(0.095)	(0.074)
Observations	2,637	2,637	2,637	2,637	2,637
R-squared	0.118	0.198	0.241	0.093	0.094
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Civic Expectations': factor variable increasing in expectations of civic behavior of others. 'Beginning': text agreement question at the beginning of the survey. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

C Empirical Appendix

C.1 Country Level Analysis

Even though treatment effects are similar across all countries, we find some differences in the average indicator of 'Health Concerns'. It is 6 in Italy and Spain, taken together (with no substantial differences between them) and 4.9 in Germany and Netherlands (with again no substantial differences between them). Yet the variance (2.85) is the same across the two pairs of countries. Trust indicators are lower in Italy and Spain compared to Germany and Netherlands, both across the entire sample as well as for non-treated respondents.[44] This is not surprising, as Proposition 2 predicts that high overall levels of concern bring about low levels of institutional and interpersonal trust.

Given these differences in initial conditions, we decided to replicate our analysis splitting the sample in two across these country lines in Figure C2, and in Tables C13 and C14 in the Appendix. What immediately stands out in the left panels of Figure C2 is that the treatment effect on social trust is more pronounced in Italy and Spain than in Germany and the Netherlands. This finding is consistent with our theory. In line with Proposition 3, the treatment may have caused all respondents to focus on the strictness of the containment rules, resulting in stronger negative effects on the interpersonal trust of those who were more concerned. Larger drops in interpersonal trust are then expected exactly in Spain and Italy, where people were more worried, than in Germany and the Netherlands.

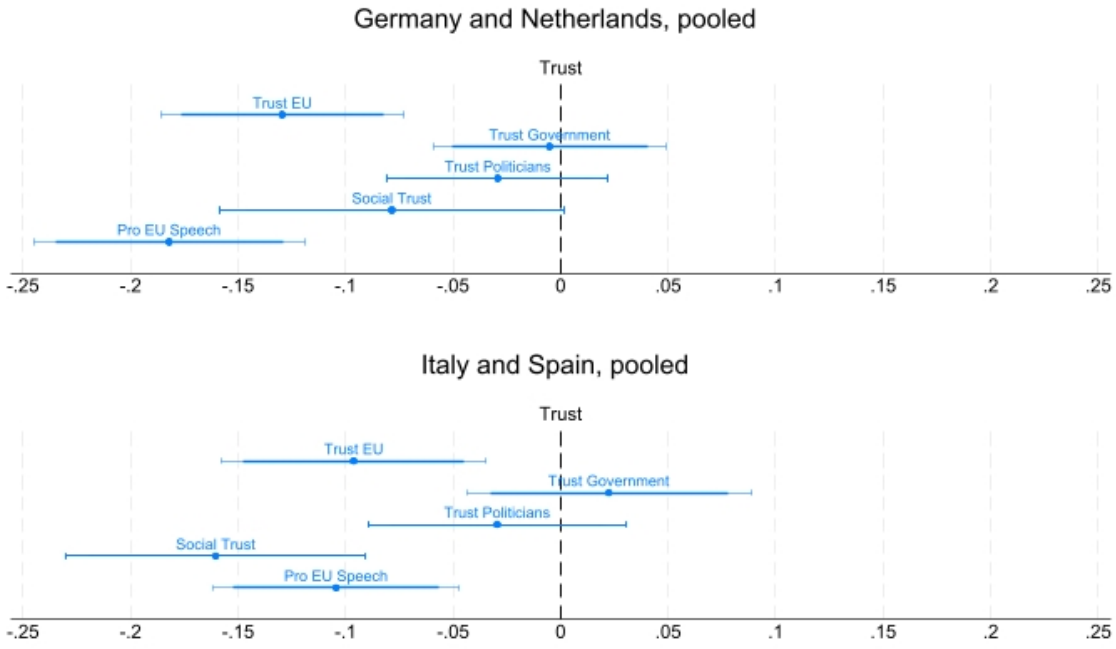
In Table C18, we lastly replicate the heterogeneity analysis for the same country subsets. The interaction effects for social and institutional trust are not substantially different across the two subsets.

C.2 Placebo Test

As mentioned above, the original survey included a wide set of outcome questions which are not related to the theoretical model. These are questions on other dimensions of trust (police and media), identity, giving up freedom in exchange for more safety, values

[44]The average measure of social trust in Italy and Spain is 4.8 (and 5.1 in the untreated group) with no significant differences between the two countries. It is 5.3 (and 5.4 in the untreated group) in Germany and Netherlands, with again no large differences between them. Trust in the government is 3.7 in Italy-Spain (3.6 in the untreated group), while it is 5.3 (and 5.3 with no treatment) in Germany and Netherlands.

Figure C2: Country Level Analysis



Note: The figures display the impact of Covidfirst on our set of outcomes for the two groups of countries, Germany-Netherlands (TOP PANEL), Italy-Spain (BOTTOM PANEL), together with 95% (delimited by vertical bars) and 90% (bold line) confidence intervals.

related to universalism and local traditions, and policy preferences related to populism (immigration, free media, having a strong leader in power, and excessive elite power). In this case, the model does not predict any heterogeneous effects depending on the level of Covid-19-related health concerns. We therefore use them as Placebo outcomes. In Figure C4 we replicate the model of Table 3 but we only report the interaction term *Covidfirst * Health Concerns.* The figure reports in all cases null effects on this set of outcomes unrelated to social trust and trust in political institutions.

C.3 Further tables for Section C.3

Table C4: Priming Covid, Trust and Health Concerns: tertiles

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Covidfirst	-0.155***	-0.071*	-0.111***	-0.069	-0.208***
	(0.038)	(0.037)	(0.033)	(0.042)	(0.039)
Covidfirst * Health Concerns 2^{nd} Tertile	0.105**	0.122**	0.128**	-0.058	0.109**
	(0.050)	(0.052)	(0.052)	(0.058)	(0.054)
Covidfirst * Health Concerns 3^{rd} Tertile	0.012	0.125**	0.124**	-0.108*	0.099**
	(0.068)	(0.063)	(0.059)	(0.056)	(0.047)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.056	0.156	0.213	0.080	0.089
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Health Concerns': perceived Covid-19 health risks from 1 (very low) to 10 (very high) split in three tertiles. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C5: Priming Covid-19, Trust, Health Concerns: by treatment

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
T = 1 (H)	-0.102***	0.028	-0.018	-0.120***	-0.108***
	(0.028)	(0.030)	(0.028)	(0.035)	(0.029)
T = 2 (H+E)	-0.185***	-0.045	-0.086***	-0.185***	-0.182***
	(0.028)	(0.029)	(0.027)	(0.034)	(0.029)
T = 3 (H+C)	-0.060**	0.037	0.010	-0.052*	-0.140***
	(0.029)	(0.026)	(0.025)	(0.032)	(0.031)
Health Concerns	0.109***	0.067***	0.069***	-0.024	0.032**
	(0.023)	(0.022)	(0.022)	(0.022)	(0.015)
T = 1 * Health Concerns	-0.022	0.024	0.022	-0.058	0.024
	(0.033)	(0.030)	(0.028)	(0.036)	(0.025)
T = 2 * Health Concerns	0.023	0.066*	0.072**	-0.048	0.029
	(0.035)	(0.034)	(0.032)	(0.030)	(0.025)
T = 3 * Health Concerns	0.020	0.085**	0.086***	-0.015	0.059**
	(0.040)	(0.035)	(0.030)	(0.031)	(0.027)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.060	0.157	0.215	0.081	0.088
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models for each sub-condition (T=1,2,3). 'Health Concerns': perceived Covid-19 health risks from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C6: Evaluations of Covid-19 Regulations Strictness

	(1) Rules too Strict	(2) Rules too Strict	(3) Rules too Strict	(4) Gov. Performance	(5) Gov. Performance	(6) Gov. Performance
Compliance	-0.703***			0.192***		
	(0.043)			(0.015)		
Health Concerns		-0.261***			0.116***	
		(0.046)			(0.016)	
Civic Expectations			-1.201***			0.335***
			(0.059)			(0.020)
Observations	8,234	8,234	2,741	8,234	8,234	2,741
R-squared	0.078	0.037	0.191	0.187	0.169	0.270
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Rules too Strict': perceived strictness of Covid-19 regulations from 1 (not at all) to 10 (a lot). 'Gov. Performance': evaluation of the national government's crisis management from 1 (very bad) to 10 (very good). 'Compliance': factor variable increasing in compliance with Covid-19 regulations. 'Civic Expectations': factor variable increasing in expectations of civic behavior of others. 'Health Concerns': self-perceived Covid-19 health risks from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

C.4 Additional tables

Table C7: Summary statistics

Variable	Mean	Std. Dev.	N
Trust EU	-0.086	1.025	7917
Trust Nat. Gov.	-0.005	1.049	7917
Trust Politicians	-0.036	1.03	7917
Trust Others	-0.091	1.05	7917
Pro EU Speech	-0.082	1.043	7917
Tax Poverty	-0.062	1.054	7917
Tax Health Exp.	-0.034	1.061	7917
Tax Unemp. Welf.	-0.036	1.043	7917
Tax Pensions	-0.038	1.051	7917
General Tax	0.028	1.013	7917
Own Tax	0.043	1.014	7917
Health Concerns	-0.001	1.008	7917
Compliance	0	1	7917
Civic Expectations	0	1	2639
EU text position	0.5	0.5	7917
Unemployed	0.067	0.25	7917
Education	3.979	1.486	7917
Italian born	0.949	0.219	7917
Female	0.5	0.5	7917
Age	45.37	13.928	7917
Household size	2.366	0.983	7917
Voted	0.861	0.346	7917
Income Adj.	1237.645	697.689	7917
Single	0.316	0.465	7917

Table C8: Social Media Usage and Trust

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Social Media	-0.172***	-0.147***	-0.095***	-0.088***	-0.052*
	(0.030)	(0.040)	(0.036)	(0.030)	(0.030)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.050	0.148	0.204	0.077	0.084
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Social Media': is a dummy equal to 1 for respondents that get informed mostly on social media (0 otherwise, the other choices were TV news, newspapers, radio/podcasts). Controls include treatment status, gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C9: Priming Covid-19 and Trust

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Covidfirst	-0.114***	0.009	-0.029	-0.120***	-0.143***
	(0.021)	(0.022)	(0.020)	(0.027)	(0.023)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.047	0.146	0.203	0.076	0.084
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C10: Priming Covid-19 and Trust by Treatment Group

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
T = 1	-0.101***	0.030	-0.017	-0.120***	-0.107***
	(0.029)	(0.030)	(0.029)	(0.035)	(0.029)
T = 2	-0.181***	-0.041	-0.082***	-0.187***	-0.180***
	(0.029)	(0.029)	(0.028)	(0.034)	(0.030)
T = 3	-0.059**	0.038	0.011	-0.053*	-0.140***
	(0.030)	(0.027)	(0.026)	(0.032)	(0.032)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.048	0.147	0.203	0.078	0.084
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models for each sub-condition (T=1,2,3). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C11: Priming Covid-19 and Tax

	(1) Tax Poverty	(2) Tax Health Exp.	(3) Tax Unemp. Welf.	(4) Tax Pensions	(5) General Tax	(6) Own Tax
Covidfirst	-0.100***	-0.072***	-0.058***	-0.072***	0.033	0.059**
	(0.023)	(0.024)	(0.022)	(0.022)	(0.020)	(0.026)
Observations	8,234	8,234	8,234	8,234	8,234	8,234
R-squared	0.029	0.059	0.064	0.052	0.134	0.117
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. Controls include gender, age groups, employment status, education, immigrant status, family status and number of family members, equivalised household income (coded into five quantiles)a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C12: Priming Covid-19 and Tax by Treatment Group

	(1) Tax Poverty	(2) Tax Health Exp.	(3) Tax Unemp. Welf.	(4) Tax Pensions	(5) General Tax	(6) Own Tax
T = 1	-0.082***	-0.048	-0.054*	-0.078***	0.011	0.017
	(0.029)	(0.033)	(0.029)	(0.029)	(0.023)	(0.030)
T = 2	-0.110***	-0.110***	-0.073**	-0.079***	0.076**	0.081**
	(0.033)	(0.030)	(0.029)	(0.029)	(0.030)	(0.039)
T = 3	-0.093***	-0.040	-0.038	-0.056*	0.025	0.076***
	(0.030)	(0.031)	(0.029)	(0.031)	(0.026)	(0.028)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.033	0.063	0.071	0.064	0.150	0.130
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models for each sub-condition (T=1,2,3). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

Table C13: Priming Covid-19 - Germany-Netherlands

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech	
Covidfirst	-0.129***	-0.005	-0.029	-0.079*	-0.182***	
	(0.029)	(0.028)	(0.026)	(0.041)	(0.032)	
Observations	4,035	4,035	4,035	4,035	4,035	
R-squared	0.050	0.062	0.055	0.099	0.070	
Controls	YES	YES	YES	YES	YES	
Region FE	YES	YES	YES	YES	YES	
	(1) Tax Poverty	(2) Tax Health Exp.	(3) Tax Unemp. Welf.	(4) Tax Pensions	(5) General Tax	(6) Own Tax
Covidfirst	-0.100***	-0.082***	-0.068**	-0.087***	0.003	0.052
	(0.034)	(0.029)	(0.033)	(0.031)	(0.027)	(0.035)
Observations	4,035	4,035	4,035	4,035	4,035	4,035
R-squared	0.032	0.018	0.024	0.041	0.032	0.035
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C14: Priming Covid-19 - Italy -Spain

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech	
Covidfirst	-0.096***	0.022	-0.030	-0.160***	-0.104***	
	(0.031)	(0.033)	(0.030)	(0.035)	(0.029)	
Observations	3,881	3,881	3,881	3,881	3,881	
R-squared	0.052	0.074	0.054	0.051	0.042	
Controls	YES	YES	YES	YES	YES	
Region FE	YES	YES	YES	YES	YES	
	(1) Tax Poverty	(2) Tax Health Exp.	(3) Tax Unemp. Welf.	(4) Tax Pensions	(5) General Tax	(6) Own Tax
Covidfirst	-0.086**	-0.049	-0.038	-0.050*	0.078**	0.068
	(0.033)	(0.037)	(0.030)	(0.029)	(0.031)	(0.041)
Observations	3,881	3,881	3,881	3,881	3,881	3,881
R-squared	0.032	0.061	0.035	0.084	0.217	0.197
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C15: Priming Covid-19, Tax and Compliance

	(1) Tax Poverty	(2) Tax Health Exp.	(3) Tax Unemp. Welf.	(4) Tax Pensions	(5) General Tax	(6) Own Tax
Covidfirst	-0.097***	-0.069***	-0.056**	-0.068***	0.034*	0.061**
	(0.023)	(0.024)	(0.023)	(0.022)	(0.020)	(0.026)
Compliance	0.130***	0.184***	0.144***	0.187***	0.065***	0.088***
	(0.021)	(0.023)	(0.023)	(0.022)	(0.020)	(0.024)
Covidfirst*Compliance	0.025	-0.016	0.003	-0.026	-0.032	-0.042*
	(0.024)	(0.025)	(0.027)	(0.025)	(0.021)	(0.024)
Observations	8,234	8,234	8,234	8,234	8,234	8,234
R-squared	0.045	0.081	0.080	0.074	0.136	0.120
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Compliance': level of compliance with Covid-19 regulations from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C16: Priming Covid-19, Trust and Compliance

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Covidfirst	-0.111***	0.011	-0.028	-0.120***	-0.142***
	(0.021)	(0.021)	(0.020)	(0.027)	(0.022)
Compliance	0.120***	0.098***	0.085***	0.015	-0.003
	(0.024)	(0.022)	(0.023)	(0.021)	(0.020)
Covidfirst * Compliance	0.021	0.067***	0.050**	-0.064***	0.045**
	(0.027)	(0.022)	(0.020)	(0.023)	(0.022)
Observations	7,916	7,916	7,916	7,916	7,916
R-squared	0.061	0.162	0.214	0.078	0.085
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Compliance': level of compliance with Covid-19 regulations from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table C17: Priming Covid-19, Trust and Civic Expectations

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Pro EU Speech
Covidfirst	-0.068*	0.036	0.004	-0.065*	-0.149***
	(0.037)	(0.032)	(0.033)	(0.039)	(0.045)
Civic Expectations	0.151***	0.192***	0.140***	0.033	-0.036
	(0.035)	(0.031)	(0.033)	(0.029)	(0.041)
Covidfirst * Civic Expectations	0.022	0.050	0.029	-0.112***	0.052
	(0.044)	(0.040)	(0.040)	(0.039)	(0.048)
Observations	2,637	2,637	2,637	2,637	2,637
R-squared	0.094	0.198	0.241	0.118	0.093
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Civic Expectations': factor variable increasing in expectations of civic behavior of others. Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

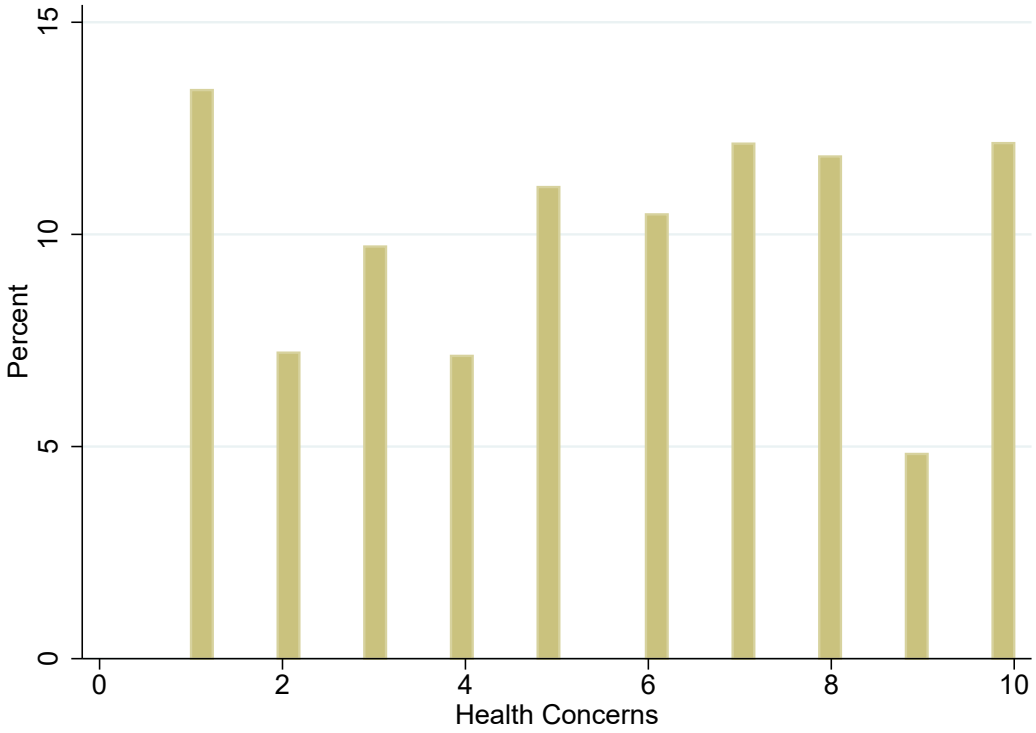
Table C18: Priming Covid-19, Trust and Health Concerns - Different countries

Italy-Spain	(1) Trust Others	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Pro EU Speech	(5) Trust EU
Covidfirst	-0.156***	0.008	-0.047	-0.114***	-0.095***
	(0.036)	(0.034)	(0.030)	(0.030)	(0.032)
Health Concerns	-0.020	0.020	0.019	0.034	0.116***
	(0.028)	(0.030)	(0.023)	(0.021)	(0.033)
Covidfirst* Health Concerns	-0.016	0.057	0.074**	0.035	-0.033
	(0.035)	(0.040)	(0.031)	(0.028)	(0.045)
Observations	3,881	3,881	3,881	3,881	3,881
R-squared	0.052	0.078	0.061	0.062	0.059
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES
Germany-Netherlands	(1) Trust Others	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Pro EU Speech	(5) Trust EU
Covidfirst	-0.088**	0.008	-0.018	-0.177***	-0.120***
	(0.042)	(0.029)	(0.027)	(0.032)	(0.028)
Health Concerns	-0.035	0.117***	0.117***	0.037*	0.111***
	(0.034)	(0.035)	(0.039)	(0.022)	(0.032)
Covidfirst*Health Concerns	-0.044	0.058*	0.050	0.023	0.040
	(0.034)	(0.034)	(0.035)	(0.028)	(0.036)
Observations	4,035	4,035	4,035	4,035	4,035
R-squared	0.103	0.087	0.077	0.072	0.069
Controls	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES

Note: Estimates from OLS models. 'Covidfirst': all treatments pooled together. 'Health Concerns': self-perceived Covid-19health risks from 1 (very low) to 10 (very high). Controls include gender, age, age squared, employment/unemployment, years of education, immigrant status, family status and number of family members, a dummy for voting in the last elections, equivalised household income (coded into five quantiles), a dummy for the position of the pro-EU text within the survey and region fixed effects. All controls are omitted to enhance readability. Robust standard errors clustered at the province level are in parentheses. * p < 0.1; ** p < 0.05; *** p < 0.01

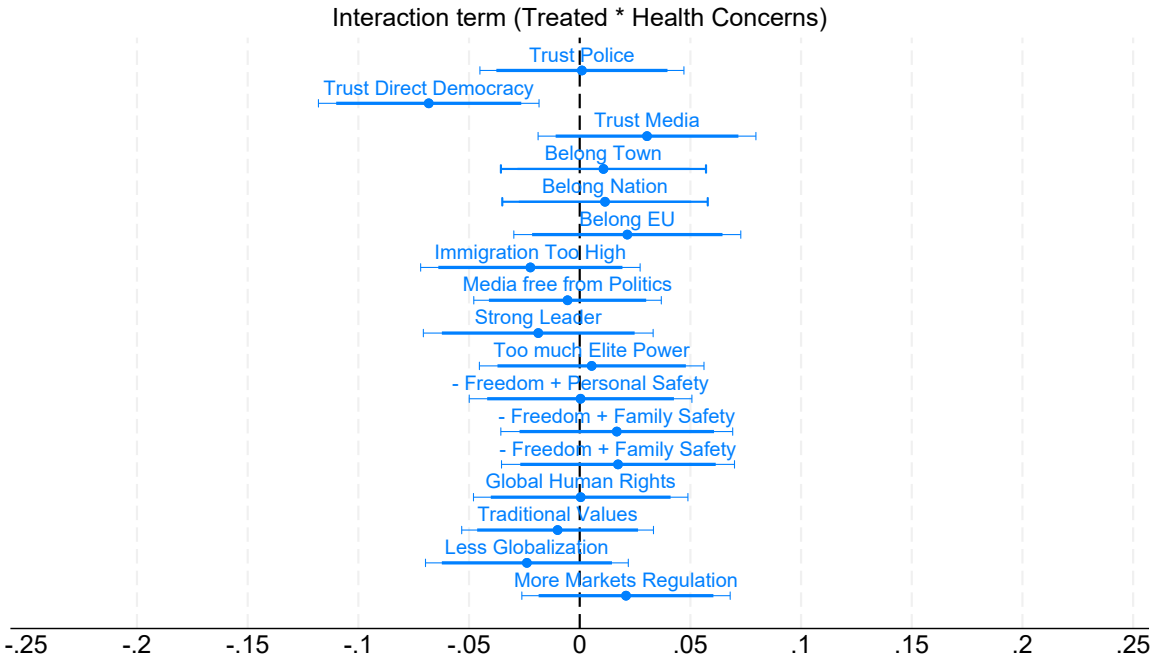
C.5 Additional figures

Figure C3: Distribution of the Health Concerns Variable



The figures displays the distribution of the variable Health Concerns.

Figure C4: Placebo: Interaction Term



The figures display the interaction term COVIDFIRST * Health Concerns on an alternative set of outcomes included in the original survey. From the top: trust in police, trust in media, attachment to your town/country/European Union, immigration is too low (0) or too high (10) in your country, media should be independent from politics, preferences for a strong leader in power, elite have too much power, give up freedom for your own/family/public safety, in favor of global human rights, respect for traditions, in favor of globalization (0) or national self-sufficiency (10), in favor of market regulation. See Appendix D for the complete list of questions in the survey.

C.6 Multiple Hypothesis & Placebo Tests

We correct the p-values of our coefficients for the multiplicity of tested hypotheses using the Romano-Wolf correction algorithm (Romano and Wolf, 2005; Clarke et al., 2020). Our results are robust to the corrections. As we show below, hold true with few exceptions.

Table C19 reports corrected p-values for the estimates in the left panel of Figure 3. Comparing those estimates with the p-values in Table C19 reveals that all our significant results remain significant after correcting for multiple outcomes. Table C20 corrects the p-values for the estimates in the right panel of Figure 3. Except for a few outcomes, the estimates remain within significance range. The outcomes for which significance is lost are Trust EU, Social Trust and Trust Direct Democracy in the H + C condition.

Table C19: Multiplicity correction for the estimates reported in Figure 3, left panel

Covidfirst	Model p-value	Romano-Wolf p-value
Trust EU	<0.001	0.002
Trust Government	0.686	0.649
Trust Politicians	0.141	0.194
Social Trust	<0.001	0.002
Pro EU Speech	<0.001	0.002

Table C20: Multiplicity correction for the estimates reported in Figure 3, right panel

H	Model p-value	Romano-Wolf p-value
Trust EU	<0.001	0.004
Trust Government	0.324	0.443
Trust Politicians	0.563	0.549
Social Trust	<0.001	0.004
Pro EU Speech	<0.001	0.004
H + E	Model p-value	Romano-Wolf p-value
Trust EU	<0.001	0.002
Trust Government	0.164	0.168
Trust Politicians	0.003	0.012
Social Trust	<0.001	0.002
Pro EU Speech	<0.001	0.002
H + C	Model p-value	Romano-Wolf p-value
Trust EU	0.048	0.176
Trust Government	0.160	0.265
Trust Politicians	0.671	0.677
Social Trust	0.095	0.239
Pro EU Speech	<0.001	0.002

Table C21 reports multiplicity corrections for table C16. Save for the coefficient on the text agreement question, all significant coefficients remain significant after correction.

Table C21: Multiplicity correction for the estimates reported in Table C16

Covidfirst	Model p-value	Romano-Wolf p-value
Trust EU	<0.001	0.002
Trust Government	0.610	0.593
Trust Politicians	0.164	0.227
Social Trust	<0.001	0.002
Pro EU Speech	<0.001	0.002
Compliance	**Model p-value**	**Romano-Wolf p-value**
Trust EU	<0.001	0.002
Trust Government	<0.001	0.002
Trust Politicians	<0.001	0.004
Social Trust	0.446	0.684
Pro EU Speech	0.962	0.961
Covidfirst X Compliance	**Model p-value**	**Romano-Wolf p-value**
Trust EU	0.484	0.463
Trust Government	0.002	0.014
Trust Politicians	0.016	0.046
Social Trust	0.06	0.034
Pro EU Speech	0.051	0.108

Finally, Table C22 corrects the estimates reported in Table 3. These estimates confirm the robustness of our findings to multiplicity, with few exceptions. We lose significance of the interaction term between Covidfirst and the degree of concern for Social Trust, trust in the government and the Pro EU-Speech question.

Table C22: Multiplicity correction for the estimates reported in Table 3

Covidfirst	Model p-value	Romano-Wolf p-value
Trust EU	<0.001	0.002
Trust Government	0.747	0.732
Trust Politicians	0.116	0.170
Social Trust	<0.001	0.001
Pro EU Speech	<0.001	0.001
Health Concerns	**Model p-value**	**Romano-Wolf p-value**
Trust EU	<0.001	0.002
Trust Government	0.003	0.022
Trust Politicians	0.002	0.018
Social Trust	0.389	0.379
Pro EU Speech	0.038	0.096
Covidfirst X Health Concerns	**Model p-value**	**Romano-Wolf p-value**
Trust EU	0.840	0.834
Trust Government	0.024	0.100
Trust Politicians	0.011	0.072
Social Trust	0.061	0.198
Pro EU Speech	0.057	0.198

C.7 Sample size and balance test

Tables C23 and C24 report respectively the sample size for each country and the balance of our sample across conditions.

Table C23: Sample size per country

Country	Sample size	Share of total
Germany	2161 obs.	26.24%
Italy	2003 obs.	24.32%
Netherlands	2071 obs.	25.15%
Spain	2000 obs.	24.29%
Total	8235 obs.	100.00%

Table C24: Balancing Test

Variables	Δ: Covidfirst-Baseline	Δ: Health-Baseline	Δ: Economic-Baseline	Δ: Conflict-Baseline
Unemployed	-0.012**	-0.012	-0.014*	-0.009
	(0.041)	(0.104)	(0.064)	(0.238)
Education	0.076**	0.083*	0.059	0.084*
	(0.029)	(0.065)	(0.187)	(0.058)
Native	-0.000	-0.002	0.003	-0.001
	(0.998)	(0.804)	(0.676)	(0.864)
Female	-0.018	-0.031**	-0.014	-0.008
	(0.127)	(0.039)	(0.355)	(0.574)
Age	-0.002	0.011	-0.009	-0.008
	(0.834)	(0.335)	(0.428)	(0.498)
Household size	0.032	0.017	0.051*	0.029
	(0.160)	(0.578)	(0.080)	(0.320)
HH income	-0.021	-0.015	-0.005	-0.042*
	(0.271)	(0.525)	(0.853)	(0.082)
Single	0.013	0.024*	0.006	0.010
	(0.233)	(0.092)	(0.692)	(0.488)
Observations	8,235	4,571	4,572	4,570

Sample balance table displaying the difference in means across conditions for all our observables, and its significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

D Links to local language surveys

The local language surveys and the English translation can be found at the links below.

English translation: http://www.gianmarcodaniele.com/quest.html

Dutch: https://taxmpg.eu.qualtrics.com/jfe/form/SV_85Ocx8lc4806tzT

German: https://taxmpg.eu.qualtrics.com/jfe/form/SV_5ouJ8nUBnj11lMp

Italian: https://taxmpg.eu.qualtrics.com/jfe/form/SV_5apXa5HwDkB55it

Spanish: https://taxmpg.eu.qualtrics.com/jfe/form/SV_0ln9O2bfxiBsH1r