

Daniele, Gianmarco; Martinangeli, Andrea F.M.; Passarelli, Francesco; Sas, Willem; Windsteiger, Lisa

Working Paper

Externalities and the Erosion of Trust

CESifo Working Paper, No. 10474

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Daniele, Gianmarco; Martinangeli, Andrea F.M.; Passarelli, Francesco; Sas, Willem; Windsteiger, Lisa (2023) : Externalities and the Erosion of Trust, CESifo Working Paper, No. 10474, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/279223>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Externalities and the Erosion of Trust

*Gianmarco Daniele, Andrea F.M. Martinangeli, Francesco Passarelli, Willem Sas,
Lisa Windsteiger*

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Externalities and the Erosion of Trust

Abstract

We present a theory linking political and social trust to explain trust erosion in modern societies. Individuals disagree on the seriousness of an externality problem, which leads to diverging policy opinions on how to solve it. This heterogeneity has two important effects on trust. First, disappointment with the policy rule enacted by the government breeds institutional distrust. Individuals that are more worried blame the government because the rule is too lenient. The less worried blame it even more because it is too intrusive. Second, as the rule also shapes individuals' notion of civic behavior, it drives a wedge between what an individual expects from others and their actual behavior. This fuels social distrust. The more individuals are worried, the more they distrust others that are not complying with the rules. Our experimental survey conducted in four European countries shows how these trust dynamics came to the surface during the Covid-19 pandemic. Once led to think intensely about the virus, lower institutional trust was reported predominantly by respondents that were less worried about the virus, whereas social trust declined (more) for worried individuals. We lastly find that support for the welfare state erodes alongside sliding trust levels.

JEL-Codes: D700, D720, H300, O520.

Keywords: social trust, institutional trust, heterogeneity, externalities, regulation, survey experiment, Covid-19, climate change, welfare, taxation.

Gianmarco Daniele
University of Milan / Italy
daniele.gianmarco@gmail.com

Andrea F.M. Martinangeli
Université Bourgogne Franche Comté / France
andrea.martinangeli@bsb-education.com

Francesco Passarelli
Bocconi University / Milan / Italy
francesco.passarelli@unibocconi.it

Willem Sas
University of Stirling / United Kingdom
willem.sas@stir.ac.uk

Lisa Windsteiger
University of Salzburg / Austria
lisa.windsteiger@plus.ac.at

May 26, 2023

We are grateful to Tito Boeri, Davide Cantoni, David Card, Doug Irwin, Guido Tabellini, and Francesco Trebbi for their helpful comments. This work was presented at several venues (Glasgow, St Andrews, Munich, Lisbon, Fürth, Rome): 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 is gratefully acknowledged. The paper was awarded the 'Bank of Portugal best paper prize' during the 2022 Lisbon Meeting on Economics and Public Science.

1 Introduction

Trust in political institutions is a fundamental ingredient for the stability of democracies and the effectiveness of government action. However, in the last decades political trust has been declining in nearly all Western democracies.¹ At the same time trust in others is also declining, fuelled by people’s suspicions that while the government faces growing difficulties implementing its policies, not everyone is willing to do their part.² This scenario of general dissatisfaction breeds political conflict and creates fertile ground for populism and political polarization, especially during severe economic and social crises (e.g. [Dustmann et al. \(2017\)](#); [Passarelli and Tabellini \(2017\)](#)). Sadly, whenever we 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.

Vast research in sociology, political science, and economics has shown 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 \(2008b\)](#), among many others). However, only a few contributions have tried to explain why advanced democracies are struggling to maintain high levels of institutional and interpersonal trust, especially in turbulent periods (e.g. [Friedman \(2005\)](#)). 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.

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 induce virtuous behavior among economic agents, in order to mitigate negative

¹According to Pew Research Center, trust in government in the US began eroding in the Sixties. From 75% of Americans saying that *they trust the government to do what is right just about always/most of the time* 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.eurofound.europa.eu/news/news-articles/trust-in-institutions-continues-to-fall-in-eu-despite-declining-unemployment-and-phasing-out-of;>

<https://www.oecd.org/governance/trust-in-government/>

²[Putnam et al. \(2000\)](#) documents a sharp decrease in social trust in US between 1960 and 2000. In July 2022, Pew Research Center reported that 71% of Americans thought that interpersonal confidence has worsened in the past 20 years. Most of them thought this is due to political polarization and lack of political trust. 49% of them thought that Americans are less reliable than they used to be. Democrats were more likely to think Americans’ level of interpersonal trust is a big problem for the country. [Sarracino \(2012\)](#) found a negative trend in social trust for Great Britain, France, Spain and Ireland between 1980 and 2005 and a positive trend in Sweden, Norway and Denmark. [Delhey et al. \(2018\)](#) show that in the 2007-16 period average interpersonal trust has decreased significantly in the EU28. [Genschel et al. \(2021\)](#) find it further decreased in the pandemic years of 2020-21.

externalities.³ At the same time, individuals have their own expectations about the rules government should adopt *and* about people’s behavior in response to government rules. If they think policies are wrong, they lose trust in political institutions. If the behavior of others falls short of their expectations, then they lose trust in others. Importantly, the two types of distrust are related because expectations about the behavior of others depend on government policies. When the rules are stricter a greater number of individuals may choose not to comply even when facing the risk of being caught, breeding more disappointment and interpersonal distrust.

Crucially then, what matters is the degree to which people disagree. We assume that individuals have different beliefs about the severity of an externality problem, and hence will have different opinions about what government should do. For the same reason, they also have different incentives to comply with government policies. This *heterogeneity* in beliefs could derive from different material interests, different information or ability to process it, or from different ideological or cultural traits. 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 entitled policies (expectations) with actual policies, and their own resulting behaviour with that of others.

In order to focus on how distrust is determined, we assume there are no other political distortions. The government makes its decisions averaging across all the different opinions in society to maximize the sum of individual utilities. On the one side, those who feel more strongly about an issue and value high individual mitigating behavior expect more restrictive policies and are disappointed because the government does too little. They are, however, more willing to abide by the (even mild) government rules, and will distrust more those who do not comply. On the other side, individuals who care less about an issue will distrust the government more for setting a policy they believe is ineffective or too intrusive. They have less incentives to comply with government rules and therefore will only distrust those who comply less than them.

Our model thus shows that heterogeneity of opinions yields *asymmetric* expressions of political and interpersonal distrust. The most concerned individuals are more likely to develop interpersonal distrust because they think the irresponsible behavior of the non-complying causes large social harm. Less concerned individuals, on the other hand, develop feelings of distrust mainly towards the institutions because they think the government is uselessly limiting their freedom. This ‘asymmetry’ 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 the government and ‘civic’ groups mostly distrusting the disobedient ones.

³See e.g. QuantGov’s database which measures and studies regulation in the US, showing that from 1970 to 2022, has almost tripled. <https://www.reghub.ai/data/custom>. Taxation would be an alternative to regulation. However, besides distortionary effects when it reaches high levels, taxing behavior is not always technically feasible.

Our empirical analysis suggests that the trust dynamics described by the model come to the surface 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 found that, conditional on being primed with several salient aspects of the crisis, individuals lost trust both in institutions and in others. However, they lost a lot more trust in institutions than in others if they indicated they were not too worried about Covid-19. The exact opposite holds for the more concerned respondents, who lost more trust in others compared to the less concerned ones. To proxy for concern in this context, we used survey questions eliciting self-perceived health risk.

As for the identification of the underlying channels of our model, we found strong negative correlations between the level of health concern and the view that regulations were too strict, indicating that the non-concerned indeed thought the rules were too intrusive or ineffective. We also found 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. This finding suggests that concerned individuals were indeed more demanding of others yet more lenient towards the government, whilst themselves following the rules. It is hence plausible that, once led to think intensely about Covid by our priming treatments, more concerned individuals are more disappointed with others because of their non-compliant behavior, while the less-concerned are more disappointed with the government because they deem the rules too intrusive.

We lastly document the implications of these effects in the realm of taxation, where support for different kinds of taxation moves parallel to trust. Our findings suggest that support for a well-functioning welfare state is set to erode alongside sliding trust levels. Support among concerned respondents drops less than for the non-concerned, echoing the asymmetric trust effects described above.

Our framework yields several additional realistic insights. First, 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 would translate into growing social tension and polarization, a result that is consistent with pro-cyclical patterns of distrust uncovered by the empirical literature ([Stevenson and Wolfers, 2011](#)).

Second, as social and political crises become more frequent, societies lose social capital, both in the form of political trust and interpersonal trust. This compounding effect would explain why trust has been declining in advanced democracies over the last decades, characterized by stagnation, social crises, and increasing complexity ([Algan et al., 2017](#)).

Third, the more opinions are diversified or polarized, the bigger the loss of social and political trust. Social media, sometimes artificially fueling the heterogeneity of opinions, would therefore play an important role in fomenting distrust and social tension ([Gentzkow, 2016](#)). Even the increase in cultural diversity, if associated with heterogeneous expectations about the government policy and about individual behavior, might have a negative effect

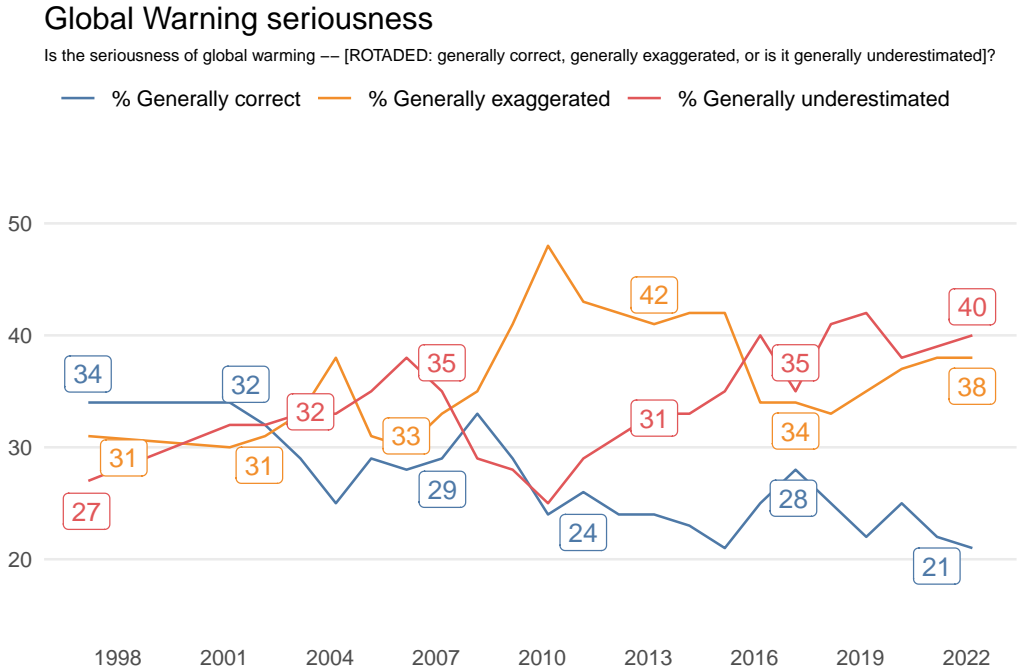
on social and political trust (Alesina and La Ferrara, 2000; Putnam, 2007). In this scenario of generalized distrust, political rhetoric emphasizing conflict rather than cooperation, or local rather than universal values would have a greater appeal to the electorate (Enke, 2020; Bonomi et al., 2021).

These insights also point at the importance of education, to enable people to consciously disentangle between fake and true information sources (Allcott and Gentzkow, 2017). There are many reasons why education is essential for the well-functioning of democracies (Glaeser et al., 2007). Our model provides another reason for why this is true.

Summing up, while one of the landmark achievements of liberal democracies is that political action draws from a wide spectrum of opinions, our framework points at the risk that when this spectrum becomes too wide, heterogeneity may become too much of a good thing. When new challenges need government intervention, disagreement on what the government should do and misaligned expectations on what people should do yield distrust. After the Covid-19 pandemic, the next big challenge nations will face is global warming. Here too our model flags the risk of a distrust crisis.

Tackling climate change might indeed require enormous government intervention and restrictions on individuals' 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. The interests at stake are enormous, and some in favor of restrictions today may not realize ex-ante how radically their behavior will have to change tomorrow.

Figure 1: Global warming seriousness



Source: GALLUP, news.gallup.com/poll/1615/environment.aspx

While PEW Research Center finds that 54% of Americans see climate change as a top threat in 2022, Gallup reports that 38% of them think the environmental threat is exaggerated, almost as much as those thinking it is correct (40%) - see Figure 1.⁴ As for Europe, the picture is scattered, with countries showing very high and very low levels of concern. Eurobarometer reports that average concerns are higher than the US, but still with a substantial share of skeptical people who are relatively unwilling to take any action to fight climate change or to support green governmental policies.⁵ Somehow interestingly, data from the New York Times shows there is clear partisan divide between concerned and unconcerned US citizens.⁶

It is apparent from these data that heterogeneity of opinions regarding climate change does exist. To some extent, it is driven by ideological or social identity elements reflected by party affiliation and may even be widening in the future (Canen et al., 2020).

Our model predicts a generalized loss of trust in this scenario. We are likely to see more sceptical ‘no-warming’ individuals in the future, angry at the government because they consider the rules too strict, similar to the ‘no-vax’ adherents during the pandemic. On the other hand, more worried people will lose trust in others, as their expectations of a joint effort to reverse or slow down global warming will be dashed. Global warming is likely to bring about a huge trust crisis exactly when unprecedented effort will be required. Recently, Besley and Persson (2022) have shown that in the absence of commitment, green policies may be too mild. Maggi and Staiger (2022) have argued that climate policies are likely to be implemented only when the point of no return has been passed, with a huge welfare loss for the delays. Our framework adds another layer to these political distortions: the lack of trust.

Related Literature

This paper is related to the literature studying the link between diversity in norms, values or preferences and political economy outcomes, such as civic conflicts or public good provision. Many papers document a negative relationship, sometimes using ethnic or cultural diversity as proxies of preference diversity (e.g. Easterly et al. (1997); La Porta et al. (1999); Alesina et al. (2004); Alesina and Ferrara (2005); Alesina et al. (2016)). Some of these authors argue that people do not like to share public goods with culturally different groups. Desmet

⁴The PEW data can be accessed here:
<https://www.pewresearch.org/global/2022/08/31/climate-change-remains-top-global-threat-across-19-country-survey/>

⁵As of April 2011, Europeans respondents ranked climate change first as the single most serious problem facing the world. Spread of infection diseases ranked second. Answers varied a lot by country, age, education, and occupation. While 78% of Europeans think it is a serious problem, 7% of them think it is not. Among the unconcerned ones, 59% say they have not taken any action to fight climate change, and 47% think that public financial support should not be given to green transition green.

⁶<https://www.nytimes.com/interactive/2020/02/20/climate/climate-change-polls.html> Partisan divide showed up also during the Covid-19 period (Allcott et al., 2020a) and seems to frequently characterize political debate

et al. (2017) find that cultural (and preference) diversity reduce civic conflict and increase public goods provision (see also Ottaviano and Peri (2006)). They also show that much of cultural diversity is observed within homogeneous groups, while differences between groups may become important when those differences are associated with a feeling of being part of separate groups. Our paper focuses on the lack of trust when people (potentially, but not necessarily, of the same ethnicity) have different preferences over policies and regulation. We remain agnostic as to where preference diversity comes from, yet we explain that when diversity leads to different compliance rates, people tend to distrust each other. Alesina et al. (2017) find large and increasing differences in values, trust, and other cultural traits within US and between and within European countries. They conjecture that sharing common policies in the EU may have bred conflicts and antagonized public opinions. Our model provides a mechanism that is consistent with their conjecture.

Ashraf and Galor (2013) find that lower cultural diversity is associated with increased homogeneity in preferences and higher levels of social trust. Lower fragmentation should then facilitate coordination and the solution of the commons problem. Ramos-Toro (2019) looks at carbon emissions. He finds that cultural homogeneity yields social cohesion and trust, leading to lower greenhouse gas emissions in wealthy countries. Our framework sheds some light on these findings. It provides a political-economy mechanism through which preference homogeneity yields more institutional and interpersonal trust, two essential ingredients to solve social coordination problems such as global warming.

Our framework shares common features with literature connecting trust and growth.⁷ Some works look at the relationship between trust, regulation, and economic performance. Algan and Cahuc (2009) explain that the fear of uncivic behavior leads people to demand for employment protection (an inefficient form of regulation) rather than unemployment subsidies. Thus, low trust leads to macroeconomic inefficiencies. Similar to them, we assume that regulation may trigger non-compliant behavior. But differently from them, civicness and trust are conceptually distinct in our paper. We implicitly assume that civicness is the distance between the government rule and individuals' actual behavior, while distrust is the *reaction* non-compliance incites in other people. Indeed, in our model civic/uncivic behavior and interpersonal trust are endogenously determined by regulation, and the latter depends on opinion heterogeneity in policy opinions. Pinotti (2012) corroborates the causal link between distrust and demand for regulation as posited by Algan and Cahuc (2009), whilst Durante et al. (2021) find that in Italian areas where civic capital is higher, compliance with Covid-19 regulation was stronger.

Aghion et al. (2010) follow the position that distrust brings about demand for regulation, but also that regulation itself breeds distrust via corruption and bribing. This complementarity may cause bad equilibria (with low trust, high regulation, and low growth) and good

⁷See, among many others, Guiso et al. (2004); Knack and Keefer (1997); Ichino and Maggi (2000) and Tabellini (2008a)

ones. [Martinangeli et al. \(2023\)](#) show that institutions leaving room for undetectable corrupt behaviour causes sharp drops in generalised social trust. Our model shares this idea that regulation yields distrust, but the mechanism is different. We point at disagreement with government rules, while we have no bribing, which in fact can reinforce the impact of regulation on distrust. Moreover, we point at opinion heterogeneity across people, which allows us to explain why advanced (good equilibrium) societies are experiencing loss of trust during turbulent periods.

By proposing a mechanism linking individuals' policy preferences to trust, this paper contributes to the extensive literature documenting a positive relationship between distrust and populist voting (for instance, [Algan et al. \(2017\)](#); [Dustmann et al. \(2017\)](#); [Boeri et al. \(2018\)](#); [Algan et al. \(2018\)](#); [Giuliano and Wacziarg \(2020\)](#); [Guiso et al. \(2020\)](#); [Daniele et al. \(2023\)](#), among many others. For a survey, see [Guriev and Papaioannou \(2022\)](#)). In most of this literature trust is exogenously given and there is no distinction between interpersonal and institutional trust. So, our contribution is twofold. First, our model shows how trust may endogenously depend on preference heterogeneity. Second, it distinguishes between interpersonal and institutional trust. The former is possibly more important to explain conflicts between social groups. The latter is perhaps more relevant when it comes to conflicts between citizens and the government.

By endogenizing political attitudes towards redistributive taxation in an experimental setting, this paper is close to [Di Tella et al. \(2021\)](#), [Alesina et al. \(2018\)](#), [Kuziemko et al. \(2015\)](#). They study the effect of trust on demand for redistribution in different contexts. Similar to their findings the measures of redistributive taxation in our survey move parallel to both interpersonal and institutional measures of trust.

Moving beyond purely monetary externalities, our model (with some technical changes) can also more broadly be interpreted as capturing "moral" externalities coming from behavior that goes against individuals' own moral values or social norms. Institutional distrust in this case would account for individuals' disappointment in government laws in more ethical or cultural fields, while interpersonal distrust would account for individuals' disappointment in others' behavior in those fields. In this light, our framework would be close to the literature studying the relationship between moral values and political behavior (e.g., [Bénabou and Tirole \(2006\)](#), [Enke \(2019\)](#), [Norris and Inglehart \(2019\)](#), [Guriev \(2018\)](#), [Fukuyama \(2018\)](#), [Bursztyn et al. \(2020\)](#)).

We share with [Passarelli and Tabellini \(2017\)](#) the idea that people compare actual outcomes with counterfactual entitled outcomes, as a kind of fair reference point. In their paper, this comparison brings about individual aggrivement and incentives to participate in street protests. In this paper we use two counterfactual outcomes to size up, on the one hand, individuals' distrust in the government, and, on the other hand, their distrust in others.

Our model also 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 focus on preference heterogeneity and behavior heterogeneity ties this paper to the extensive literature studying the link between information sources or narratives, political preferences (DellaVigna and Gentzkow, 2010; Bordalo et al., 2020; Bursztyn et al., 2022) and behavior (DellaVigna and Kaplan, 2007; Allcott et al., 2020a,b).

Finally, we also contribute to the literature studying the effect of (localised) epidemics on institutional trust and political preferences. Aksoy et al. (2020) find that epidemic exposure in an individual’s impressionable years has a persistent negative effect on confidence in political institutions and leaders. Foremny et al. (2020) implement an information experiment on the Covid-19 fatality rate, suggesting that preferences for health care expenditures have almost doubled. Similarly, Rees-Jones et al. (2022) find that demand for social safety nets have increased since the outbreak of the epidemic.⁸ Bor et al. (2022) find that vaccinated people express discriminatory attitudes towards the non-vaccinated, likewise towards immigrants and other minority targets. These results are consistent with our theoretical predictions and our experimental evidence.

2 A General Model of Externalities and Trust

Our model follows the tradition of the political economy literature on redistributive fiscal policy, but we focus on rules and externalities. Consider a society with a continuum of individuals. Let us normalise the population size to unity, and assume each individual is assigned a position i in the interval $[0, 1]$ following, for simplicity, a uniform distribution. Let all individuals’ positions be common knowledge. Later, it will become clear which individual trait is captured by those positions. Sometimes it will be useful to refer to “individual $i \in [0, 1]$ ” or “type 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 helps reducing emissions.

Thus we are thinking of a ‘virtuous’ dimension in people’s behavior which, besides

⁸For an overview of the rapidly expanding work on Covid-19 economic effects see Brodeur et al. (2020) and the cited literature.

private benefits, mitigates a negative externality (or prompts a positive externality).⁹ So any individual i will benefit from her own behavior, b_i , and from the mitigating behaviour of others. To capture this idea, let her benefit function be

$$\varepsilon_i(b_i + B_{-i}), \quad (1)$$

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

$$B_{-i} \equiv \int_0^1 b_j dj, \quad \forall j \neq i, \quad (2)$$

and $\varepsilon_i(\cdot)$ a twice differentiable function with $\varepsilon'_i(\cdot) > 0$ and $\varepsilon''_i(\cdot) \leq 0$.¹⁰

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

We capture concern with 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 .¹¹ 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.¹² 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, which is expressed by an increasing cost function, $c(b_i)$, e.g. the time cost of having to find and put on a mask, the effort of keeping

⁹Mitigating a negative externality or prompting a positive one are formally equivalent in this model. Hereafter, by thinking of the negative externality case, we will often refer to increasing behavior as “mitigating behavior”.

¹⁰For simplicity, 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} . Of course, more realistically i might be affected more by the behavior of some individuals (e.g. her neighbors or her relatives) than others, and with different weights. Qualitatively, however, the message of this model would not change substantially if we were to take into account different degrees of ‘proximity’ to other individuals. This model can also capture those situations in which individuals’ behavior yields pure externalities with no private gains. In this case the benefit function would be $\varepsilon_i(B_{-i})$. Note that concavity of the benefit function captures the idea that my incentive to mitigate behavior is lower at the margin when people are already mitigating theirs, as in the case of social-distancing in a pandemic. One might also assume convexity, by thinking for instance that 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 then more options are available in the train timetable. It will become clear that concavity (convexity) implies strategic substitutability (complementarity).

¹¹Here we are assuming that if $\varepsilon'_i(y) > \varepsilon'_j(y)$ holds for some y it also holds for all y , a sort of ‘concern consistency’.

¹²As 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.

social distance, the cost of losing business opportunities because of lockdown, but also the material cost of a solar panel, the cost of having to sell an old car and buy a new less-polluting one, etc. We assume the cost function is convex and twice differentiable. Since our aim is to study how individuals' preferences and choices are driven by their *perceptions* of the private and social value of mitigating behavior, we assume costs are identical 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.¹³ For simplicity, we also assume that $c'(0) = 0$.

Combining benefits and costs, individual utility then becomes

$$U_i(b_i, B_{-i}) = \varepsilon_i(b_i + B_{-i}) - c(b_i). \quad (3)$$

For all i , optimal *decentralised* behaviour maximizes $U_i(b_i, B_{-i})$ by solving the following FOC

$$\varepsilon'_i(b_i + B_{-i}) = c'(b_i), \quad (4)$$

which has an interior solution as $\varepsilon''_i(\cdot) - c''(\cdot) < 0$. Equation (4) defines i 's best response function. Note that, as long as $\varepsilon''_i(\cdot)$ 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. Utility in a decentralized equilibrium becomes

$$U_i(b_i^*, B_{-i}^*) = \varepsilon_i(b_i^* + B_{-i}^*) - c(b_i^*), \quad (5)$$

where b_i^* is i 's decentralized equilibrium behavior, and $B_{-i}^* \equiv \int_0^1 b_j^* dj$, $\forall j \neq i$.¹⁴ 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.¹⁵ This leaves scope for policy intervention to fix the externality problem.

By implicitly differentiating (4) it follows that more concerned types will engage in more mitigating behavior, $\frac{\partial b_i^*}{\partial \varepsilon_i} > 0$. This result does not derive from any form of social or moral concern, but is entirely driven by stronger private concerns. In other words, in a

¹³Alternatively one could assume that individuals are different in the effort they have to make to implement mitigating behavior. Heterogeneity would come from the cost side. Such a model would capture different situations, but it would lead to similar predictions regarding the political disagreement across people.

¹⁴Of 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 in the paper.

¹⁵Namely, she would maximize $U_i(b_i, B_{-i}) + \int_0^1 U_j(b_j, B_{-j})$ wrt b_i . The FOC would be $\varepsilon'_i(\cdot) + \int_0^1 \varepsilon'_j(\cdot) \frac{\partial B_{-j}}{\partial b_i} dj = c'(\cdot)$. Since $\int_0^1 \varepsilon'_j(\cdot) \frac{\partial B_{-j}}{\partial b_i} dj = \int_0^1 \varepsilon'_j(\cdot) dj > 0$, it is immediately apparent that the internalization equilibrium would imply a higher mitigating behavior compared to the decentralized equilibrium.

decentralized equilibrium individuals behave more virtuously only because they are more concerned about the effect of their behavior on their own welfare. In this model individuals are not altruistic, as we want to abstract away from any moral motivation for pro-social behavior.¹⁶

2.1 Behavioral Rules and Policy Entitlements

To incentivise people to engage in socially beneficial behavior, we consider a rule fixing a lower bound ρ to the behavior of all individuals.¹⁷ Individuals have different perceptions about the benefits of their behavior and the behaviors of others, so they will have different views about what the right rule should be. Let ρ_i^* be individual i 's most preferred rule. She chooses ρ_i^* assuming that individuals will fully comply with the rule if their best reply to the rule is below the rule (i.e. $b_j(\rho_i^*) = \rho_i^*$ if $b_j^*(\rho_i^*) < \rho_i^*$), while people whose best reply behavior to the rule is above ρ_i^* will simply choose their optimal behaviour $b_j^*(\rho_i^*) \geq \rho_i^*$.

In other words she assumes the rule splits society in two groups: those whose optimal behavior given the rule would be below the rule and who thus have to adapt their behavior to comply with the rule (the *affected* individuals, with $b_j^*(\rho_i^*) < \rho_i^*$) and those whose optimal behavior (given the rule) is above the rule (the *unaffected* individuals, with $b_j^*(\rho_i^*) \geq \rho_i^*$) and whose optimal behaviour in presence of the rule is thus higher than prescribed by the rule.

Now, i will choose her most preferred rule by maximizing the following indirect utility function and assuming all individuals choose their equilibrium behavior given the rule,

$$V_i(\rho) = \varepsilon_i (b_i^*(\rho) + B_{-i}^*(\rho)) - c(b_i^*(\rho)), \quad (6)$$

where

$$B_{-i}^*(\rho) = \int_0^{x(\rho)} \rho dj + \int_{x(\rho)}^1 b_j^*(\rho) dj \quad \forall j \neq i. \quad (7)$$

The first term on the RHS of (7) is the cumulative behavior of all the affected individuals forced to follow the rule, including the last affected individual, $x(\rho)$. The second term is the cumulative behavior of the non-affected individuals. The latter choose their behavior as they would in a decentralized equilibrium. But the existence of a rule yields an important difference. The rule forces the affected individuals to increase their behavior. This in turn lowers the unaffected individuals' benefit of increasing their behavior at the margin, leading some of them to reduce their behavior in equilibrium. The reason is that, as long as the benefit function is strictly concave, there is strategic substitutability. Thus, a rule

¹⁶It is easy to see that, if there were no 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.

¹⁷Of 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.

indirectly (and negatively) also affects the equilibrium behavior of those who do not have to comply.¹⁸ However if the number of affected individuals is sufficiently high, a rule increases average behavior in society, despite some non-affected individuals reducing their behavior in equilibrium.¹⁹

The optimality condition which pins down i 's bliss point, ρ_i^* , depends on whether i is affected or not by the rule. If i is affected her behavior is ρ . So the optimality condition is the following

$$\varepsilon'_i(\rho_i^* + B_{-i}^*(\rho_i^*)) \cdot (1 + x(\rho_i^*)) \leq c'(\rho_i^*), \quad (8)$$

with strict inequality implying $\rho_i^* = b_i^*(\rho_i^*)$.

If i is unaffected, then her equilibrium behavior, $b_i^*(\rho_i^*)$, is only indirectly affected by the rule. By strategic substitutability, $\frac{\partial b_i^*}{\partial \rho} < 0$. So the optimality condition is

$$\varepsilon'_i(b_i^*(\rho_i^*) + B_{-i}^*(\rho_i^*)) \cdot \left(\frac{\partial b_i^*(\rho_i^*)}{\partial \rho} + x(\rho_i^*) \right) \geq c'(b_i^*(\rho_i^*)) \frac{\partial b_i^*(\rho_i^*)}{\partial \rho} \quad (9)$$

and strict inequality also implies that $\rho_i^* = b_i^*(\rho_i^*)$.

An individual chooses a rule that is stricter than her decentralized behavior when that rule can affect a sufficiently high number of individuals.²⁰ (8) shows, this is more likely to happen when $x(\rho)$ is higher. Individual i trades the marginal benefits of increasing her own behavior and the behavior of others against the additional cost due to complying with the rule. Marginal benefits are twofold. First, those coming from her own more virtuous behavior. Second, those coming from drawing more individuals into the affected group, up to individual $x(\rho)$. Both sides of (9) are negative if it holds with equality. This means that an individual choosing a rule that is below her equilibrium behavior does it with the intent to force the affected ones to adopt more virtuous behavior and save on her private costs.

Using (8), individual i 's most preferred rule is bounded when the number of individuals who become affected by marginally increasing the rule, $x'(\rho)$, is not too high. Then, the convexity of costs and concavity of ε is sufficient to ensure concavity at ρ_i^* ,

$$\varepsilon''_i(\rho_i^* + B_{-i}^*(\rho_i^*)) (1 + x(\rho_i^*))^2 + \varepsilon'_i(\rho_i^* + B_{-i}^*(\rho_i^*)) \cdot x'(\rho_i^*) < c''(\rho_i^*), \quad (10)$$

so that each i has a uniquely preferred rule ρ_i^* .

¹⁸This is the reason why $b_i^*(.)$ and $b_j^*(.)$ in (6) and (7) depend on ρ .

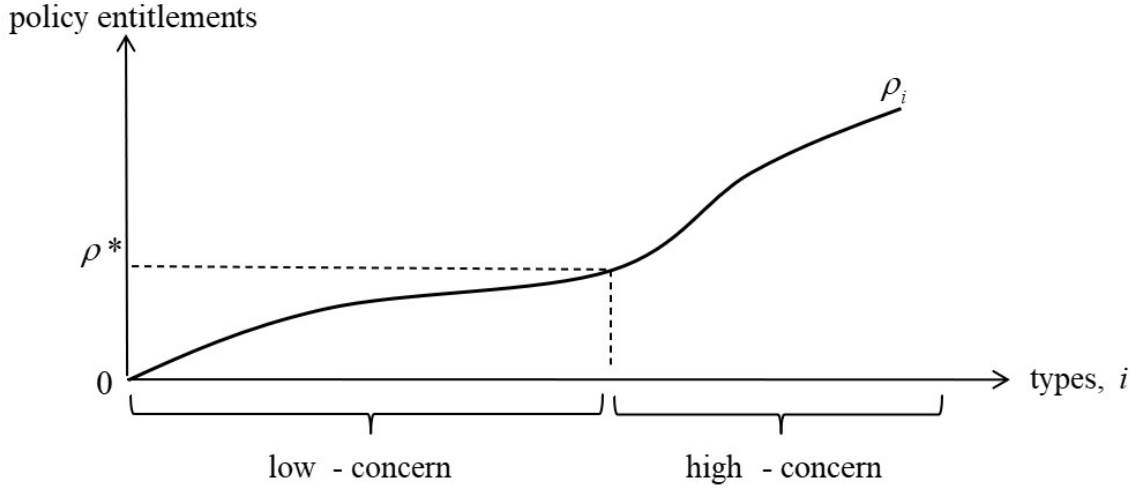
¹⁹Note that $\frac{\partial B_{-i}^*(\rho)}{\partial \rho} = x(\rho) + \int_{x(\rho)}^1 \frac{\partial b_j^*(\rho)}{\partial \rho}$. This expression represents the change in total equilibrium behavior when the rule changes at the margin. The first term on the RHS is the marginal cumulative behavior of the affected individuals. It is always positive. The second term is the marginal change of total equilibrium behavior of the non-affected ones. If it was positive, it would mean that total behavior is increasing in equilibrium, thus all non-affected individuals would have an incentive to reduce their behavior. This would yield a contradiction. Thus the second term must be negative. However, if the number of affected individuals is sufficiently large (i.e. $x(\rho)$ is sufficiently high), the total expression is positive. In words, a higher rule yields more mitigating behavior in society.

²⁰This result echoes Gitmez et al. (2020). They show that individuals have an incentive to vote for a rule that is stricter than their behavior with the intent of inducing more mitigating behavior in other people.

Let us dwell a bit on the normative meaning of ρ_i^* . One can see behavior as an individual's contribution to social welfare. Thus, from i 's point of view any individual should contribute at least ρ_i^* . Individuals are welcome to increase their behavior even further, but in i 's view no one should contribute less than ρ_i^* .²¹ The normative implication is that this rule reflects i 's 'moral' belief that individuals should not be less concerned than she is, and they should care at least as much as she cares about the consequences of their behavior on others. For these reasons we will sometimes refer to ρ_i^* as i 's 'policy entitlement', or her 'subjectively fair' rule.

By implicit differentiating (8) or (9), 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 (see Figure 2).

Figure 2: Behavioural rules & entitlements



2.2 The Socially Optimal Rule

While individuals have different views about the externality caused by individual behavior, we assume the government remains agnostic about the exact magnitude of the overall externality. It will hence maximize a standard Benthamite social welfare function to set

²¹Note that ρ_i^* is also the rule that maximizes a "distorted" Benthamite welfare function under the assumption that individuals cannot be less concerned than individual i .

policies,²²

$$W = \int_0^1 U_i(b_i) di \quad (11)$$

Differently from individuals, the government also faces the policy implementation problem. We assume it sets the policy in two steps. First, it derives the socially optimal rule, ρ^* . Second, it chooses how to enforce that rule.²³ Let us start with the first step. Similar to the individual maximization problem, once ρ^* has been set up society is split in two groups. The affected individuals below $x(\rho^*)$, and the unaffected group that voluntarily choose behaviors that are more virtuous than ρ^* . The government anticipates individuals' equilibrium behaviors, $b_i^*(\rho)$, once the rule ρ^* is enforced. The indirect social welfare function to be maximized is then,

$$W(\rho) = \int_0^1 U_i(b_i^*(\rho)) di = \int_0^{x(\rho)} (\varepsilon_i(\rho + B_{-i}^*(\rho)) - c(\rho)) di + \int_{x(\rho)}^1 (\varepsilon_i(b_i^*(\rho) + B_{-i}^*(\rho)) - c(b_i^*(\rho))) di \quad (12)$$

where $B_{-i}^*(\rho)$ is defined by (7). Then ρ^* solves the following FOC²⁴

$$\int_0^{x(\rho)} (\varepsilon'_i(\rho + B_{-i}^*(\rho)) \cdot (1 + x(\rho))) di + \int_{x(\rho)}^1 \varepsilon'_i(b_i^*(\rho) + B_{-i}^*(\rho)) \cdot \left(\frac{\partial b_i^*}{\partial \rho} + x(\rho)\right) di = \int_0^{x(\rho)} c'(\rho) di + \int_{x(\rho)}^1 c'(b_i^*(\rho)) \frac{\partial b_i^*}{\partial \rho} di \quad (13)$$

Note that the socially optimal rule equates average marginal externality on the LHS of (13) to the average marginal cost on the RHS. So government's policy reflects average concern in society. The higher the concern on average, the stricter the government's rule. Quite intuitively, that rule will look too strict to a low-concerned individual and too lenient to a high-concerned citizen. We will come back to this later.

Once ρ^* has been defined, the government has to decide on how to enforce the rule. In this context, it seems realistic that the government relies on fines and monitoring activities. It might aim at full compliance, which would be ensured by sufficiently high fines and strict

²²As mentioned earlier, we make this assumption because we want to abstract away from other political distortions. One might alternatively imagine that the government holds its own beliefs about the true benefits of mitigating behavior. Then it maximizes a paternalistic social welfare function in which individuals' utilities are the same, with the same $\varepsilon(\cdot)$. Such a model would deliver qualitatively similar predictions as the present model, as long as the government's concern lies somehow in the middle of the distribution of individuals' concerns. Our model could be extended to include elections, lobbies, protests, and other forms of political pressure. They would enter the model by distorting the government's objective function. All of these extensions would deliver interesting insights, but would probably require specific and detailed treatments.

²³Nothing really important hinges on this assumption. It simplifies the math a bit and it makes the comparison between the government's rule and individuals' entitled rules more transparent. It also captures those, perhaps realistic, situations where the enforcement problem is solved after a policy has been independently chosen. However, a model where the government simultaneously decides the rule and the enforcement instrument would yield qualitatively the same results as the present one.

²⁴By $c'(0) = 0$ it follows that the socially marginal rule is always larger than zero, so we do not have to worry about corner solutions for ρ^* . Convexity of $c(\cdot)$ and concavity of all $\varepsilon_i(\cdot)$ ensure that the stationary point is a maximum.

monitoring. This might be difficult to achieve in practice, however, especially if monitoring individuals' behavior is costly and overly high fines are considered unacceptable by society for some reason.

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

$$m \frac{(\rho^* - b_i)^+}{\rho^*} \quad (14)$$

Observe that this probability ranges from zero to one, and increases linearly with misbehavior $(\rho^* - b_i)$, at a rate $\frac{m}{\rho^*}$.

Now we can study how individuals optimally react to the government's fine and monitoring. By (14), the expected fine is $\bar{k} \cdot m \frac{\rho^* - b_i}{\rho^*}$ when behavior is $b_i < \rho^*$, while it is zero if $b_i \geq \rho^*$. Thus, individual i 's expected utility is

$$U_i(b_i, \rho^*, m) = \varepsilon_i(b_i + B_{-i}) - c(b_i) - \bar{k} \cdot m \frac{(\rho^* - b_i)^+}{\rho^*}. \quad (15)$$

In equilibrium i chooses her behavior so as to maximize (15), 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 , equilibrium behavior, $b_i^*(\rho^*, m)$, satisfies the following optimality condition at the equilibrium point:

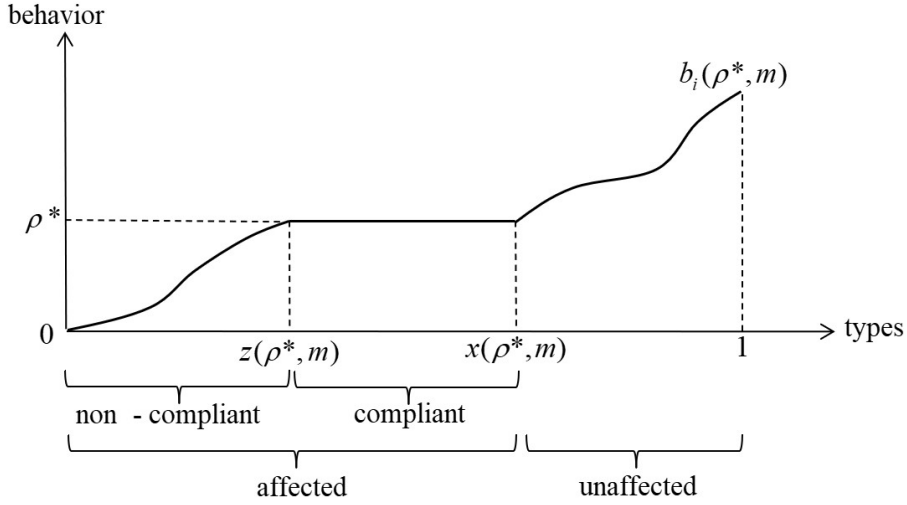
$$\varepsilon'_i(b_i^*(\rho^*, m) + B_{-i}^*(\rho^*, m)) - c'(b_i^*(\rho^*, m)) = 0 \quad \text{if } b_i^*(\rho^*, m) > \rho^* \quad (16)$$

$$\varepsilon'_i(b_i^*(\rho^*, m) + B_{-i}^*(\rho^*, m)) - c'(b_i^*(\rho^*, m)) + \frac{\bar{k} \cdot m}{\rho^*} \geq 0 \quad \text{if } b_i^*(\rho^*, m) \leq \rho^* \quad (17)$$

Equation (16) 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 3. 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 because in equilibrium they optimally react to the behavior of the affected ones, which in turn depends on m . Lemma 1 shows 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.

Expression (17) refers to the *affected* individuals. Somehow interestingly, they split in

Figure 3: Equilibrium behaviour



two sub-groups. First, those who choose to fully obey the rule and pay no fine whatsoever. They are less concerned than the unaffected ones, and more concerned than the second sub-group (see below). We know from Figure 3 that their full-compliant behavior is ρ^* and their types are higher than $z(\rho^*, m)$ and lower than $x(\rho^*, m)$. Lemma 1 also shows that the size of this group enlarges when monitoring is intensified. The second sub-group 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 with no monitoring. The perspective of being fined if caught, leads them to increase their behavior with the intent of reducing the expected fine. Lemma 1 summarizes these results.²⁵

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 $i \in ((x(.), 1]$.*
2. *Affected individuals deciding to fully comply, with $b_i^*(\rho^*, m) = \rho^*$, and $i \in [z(.), x(.)]$. The highest individual in this group, $x(\rho^*, m)$, positively depends on m and ρ^* . The lowest one, $z(\rho^*, m)$, negatively depends on m .*
3. *Affected individuals choosing not to fully comply, with $b_i^*(\rho^*, m) < \rho^*$, and $i \in [0, z(.))$. Stricter monitoring will induce more virtuous behavior by these individuals.*

The second step in the government's problem consists in choosing monitoring activity m optimally, given ρ^* and given individuals' best responses defined by Lemma 1. Socially

²⁵proofs of Lemma 1 and all following propositions are provided in Appendix B

optimal monitoring, m^* , maximizes the following indirect social welfare function:

$$\begin{aligned}
W(\rho^*, m) = & \int_0^{z(\rho^*, m)} U_i(b_i^*(\rho^*, m), B_{-i}^*(\rho^*, m)) di + \int_{z(\rho^*, m)}^{x(\rho^*, m)} U_i(\rho^*, m, B_{-i}^*(\rho^*, m)) di \\
& + \int_{x(\rho^*, m)}^1 U_i(b_i^*(\rho^*, m), B_{-i}^*(\rho^*, m)) di - k(m).
\end{aligned} \tag{18}$$

where the three integrals are aggregate utilities of the three groups defined by (15) and $k(m)$ is the monitoring cost function, which we assume is increasing, convex, and twice differentiable. To get rid, for simplicity, of redistributive and double-dividend concerns, here we also assume that fines proceeds are burnt.

The optimality condition that pins down m^* is the following

$$\begin{aligned}
\frac{\partial W}{\partial m} = & \int_0^{z(\rho^*, m)} \varepsilon'_i(b_i^*(\cdot) + B_{-i}^*(\cdot)) \frac{\partial b_i^*}{\partial m} di + \int_{x(\rho^*, m)}^1 \varepsilon'_i(b_i^*(\rho) + B_{-i}^*(\rho)) \frac{\partial b_i^*}{\partial m} di \\
\leq & k'(m) + \int_0^{z(\rho^*, m)} c'(b_i^*(\rho^*, m)) \frac{\partial b_i^*}{\partial m} di + \int_{x(\rho^*, m)}^1 c'(b_i^*(\rho^*, m)) \frac{\partial b_i^*}{\partial m} di.
\end{aligned} \tag{19}$$

Strict inequalities yield corner solutions. The first integral is the *positive* social marginal benefit of non-compliant individuals in the affected group. By Lemma 1, they increase their behavior because stricter monitoring implies a higher chance to be caught and fined. The second integral is the *negative* social marginal benefit of people in the unaffected group. They decrease their behavior because of strategic substitutability. Lemma 1 also shows that monitoring affects the size of the three groups. But these size effects are offset at the margin.

2.3 Institutional and Social Distrust

This subsection derives loss of institutional and social trust from individual expectations of a right policy and fair behavior. Individuals have different views of the beneficial effects of mitigating behavior. So each individual derives her own subjective view of what the right rule should be. We called this an ‘entitled policy’, ρ_i^* . It captures i ’s moral expectation that everyone should be at least as concerned as she is and should care about others at least as much as she does.

Institutional trust derives from i ’s expectation that the government will implement what she deems the right policy, ρ_i^* . If it does not, individual i loses trust in political institutions. Social trust reflects the confidence that other people will do what they ought to do, which in a democracy is their civic behavior of abiding by the government rules. If they do not obey the rules, our individual will start distrusting them.

In this formulation, social and institutional trust are two complementary elements of an individual’s moral view on society. On the one hand, a government is trustworthy when it implements policies reflecting her moral expectations. On the other hand, citizens are

trustworthy when they do what every civic citizen should do in a democracy: abide by the government's rules. We lastly rule out moral hypocrisy, assuming that when an individual is not complying, she can only distrust people whose behavior is worse than her own behavior.

Of course one might think of different definitions of social and institutional trust. Here however we are thinking of two concepts allowing for conceptual decomposition of individuals' disappointment. On the one hand, towards the government and, on the other, towards other citizens. Alternative formulations would imply conceptual issues. For instance, assuming that an individual distrusts others if their behavior is below her entitled policy would imply that she blames others for the same reasons why she blames the government. So interpersonal and institutional trust would be conceptually equivalent.

2.3.1 Institutional Distrust

An individual feels entitled to a level of utility corresponding to her entitled policy, ρ_i^* . As in [Passarelli and Tabellini \(2017\)](#), this entitled level of utility, denoted $\hat{V}_i(\rho_i^*)$, serves as a reference point. If utility deriving from the government policy, $V_i(\rho^*)$, falls short of that reference point, the individual is disappointed and she becomes distrustful towards the government.

Definition 1. Let D_i^I denote i 's institutional distrust. We assume it is commensurate to utility loss due to adopting ρ^* instead of ρ_i^* . With some abuse of notation, we then obtain for all i and all $j \neq i$,

$$D_i^I(\rho_i^*, \rho^*) = \hat{V}_i(\rho_i^* |_{b_i^*, b_j^* \geq \rho_i^*}) - V_i(\rho^* |_{b_i^*, b_j^* \geq \rho^*}) \quad (20)$$

Importantly, both utility levels are computed under the expectation that all individuals, including herself, will obey the rule. Under this assumption, the loss of utility is entirely due to the government. Thus, equilibrium behaviors of others, $B_{-i}^*(\rho^*)$ and $B_{-i}^*(\rho_i^*)$, are defined by (7) when the rule is ρ^* or ρ_i^* , respectively. As of individual i , equilibrium behavior, $b_i^*(\rho_i^*)$ and $b_i^*(\rho^*)$, cannot be lower than the rule.

2.3.2 Social Distrust

If loss of utility is due to low compliance, individual i cannot blame the government but will instead blame non-complying individuals. The uncivic behavior of the latter causes a loss of utility compared to the counterfactual situation in which everyone complies. If she does not comply herself, however, she can only blame people whose behavior is worse than her own behavior. Given a government rule ρ^* , the reference utility for social trust $\hat{V}_i(\rho^*, m^*)$ can thus be twofold. When individual i follows the rule or behaves even better ($b_i^* \geq \rho^*$), it is the counterfactual utility resulting from a situation where everyone else complies with that rule. When individual i does not comply with the rule ($b_i^* < \rho^*$), it is the counterfactual

utility where everyone else manages *at least* the same, insufficient, level of mitigation, b_i^* .

Definition 2. *Social distrust, D_i^S , is the loss of utility due to low compliance. For all i and all $j \neq i$, it is commensurate to the difference between i 's reference utility, $\hat{V}_i(\rho^*, m^*)$, and her actual utility in equilibrium, so that, with some abuse of notation,*

$$D_i^S(\rho^*, m^*) = \begin{cases} \hat{V}_i(\rho^*, m^*)|_{b_j \geq \rho^*} - V_i(\rho^*, m^*)|_{b_j^*, b_i^*} & \text{if } b_i^* \geq \rho^* \\ \hat{V}_i(\rho^*, m^*)|_{b_j \geq b_i^*} - V_i(\rho^*, m^*)|_{b_j^*, b_i^*} & \text{if } b_i^* < \rho^* \end{cases} \quad (21)$$

2.4 Equilibrium

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

Definition 3. *A political equilibrium consists of:*

- i. A set of entitled policies, ρ_i^* , maximizing individuals' indirect utilities defined by (6) and taking into account the equilibrium behavior of the other individuals under the assumption that ρ_i^* is fully enforced.*
- ii. A policy vector (ρ^*, m^*) such that: a) the rule ρ^* maximizes the Benthamite social welfare function defined by (11) under the assumption that all individuals comply with the rule; b) the monitoring activity m^* maximizes (18) taking into account individuals' equilibrium behavior, $b_i^*(\rho^*, m)$.*
- iii. A set of individual equilibrium behavioral responses, $b_i^*(\rho^*, m^*)$ defined by Lemma 1, such that society is split in three groups as follows: 1. a group of unaffected individuals with $b_i^*(.) > \rho^*$ and $i > x(.)$; 2.1 a group of affected individuals deciding to comply with $b_i^*(.) = \rho^*$, and $i \in [z(.), x(.)]$; 2.2 a group of affected individuals deciding not to comply, with $b_i^*(.) < \rho^*$ and $i \in [0, z(.)]$.*
- iv. A set of individual reference utilities, $\hat{V}_i(\rho^*|_{b_j \geq b_i^*})$, based on the presumption that all individuals comply with the entitled policy, ρ_i^* ; a set of counterfactual utilities, $\hat{V}_i(\rho^*, m^*|_{b_j \geq b_i^*})$, based on the presumption that all individuals comply with the government rule ρ^* or they behave at least as much as individual i does.*
- v. A set of equilibrium levels of institutional distrust, $D_i^I(\rho_i^*, \rho^*)$, defined by (20), and a set of equilibrium levels of social distrust, $D_i^S(\rho_i^*, \rho^*)$, defined by (21).*

This equilibrium features a society where, roughly speaking, the government implements a rule reflecting an average level of concern. Monitoring is costly, so it is not perfect. Some low-concern individuals choose not to fully comply. Citizens distrust the government for not implementing what they deem the right rule, and they distrust other people for not even complying with the government rule, or for complying less than they are. Below we further characterize these two forms of distrust.

Let us first point out that here we are taking a fully positive approach: we are studying what determines distrust, when the government only aims at maximizing individuals' material utility. As we discussed in the introduction, distrust can play a role in many social and political circumstances. A normative approach would imply studying the consequences of distrust, with a government internalizing them into a broader objective function. For instance, one might think of a government willing to minimize political or economic instability deriving from distrust. Internalizing how policies affect distrust might help the government shape a better policy. In this light our model can also contribute to the (possibly normative) analysis of a wide set of economic and political situations following from a rise in distrust. We will come back to this point in the conclusions.

In the society featured so far institutional distrust is triggered by different reactions to the government's policy. Quite intuitively, a highly concerned individual has a high appreciation of the benefits of mitigating behavior. So she blames the government because she thinks the government rule is too lenient and it does not allow to fully capture those benefits. A low-concern individual distrusts the government for a different reason. In her view, externalities are low. So she blames the government for having to pay an adjustment cost she thinks is not worth the low benefits of mitigating behavior. The low-concern ones are then annoyed for paying a cost to reduce an externality that they value very little. The high-concern ones are annoyed because the rule is too lenient. However, the latter are partly compensated because they can save on adjustment costs, which can be large due to cost convexity. For this reason we find that high-concern individuals are less disappointed with the government compared to the low-concern ones.

Social distrust comes from disappointment with the behavior of other people. A high-concern individual is disappointed because some people do not even abide by the government rule. In her view, their misbehavior leads to large losses compared to what they ought to do. Thus we expect high-concern individuals to develop strong feelings of social distrust. On the contrary, a low-concern individual does not think misbehavior is a big loss. Moreover, if she does not comply she can only complain for the (small) loss due to the misbehavior of those who are complying less than herself.²⁶

In sum, compared to high-concern individuals, the low-concern ones are expected to develop a stronger sense of distrust towards the government for not implementing the right policy. The high-concern ones are instead more likely to distrust other people for not abiding by the rule. These results are summarized formally by Proposition 1 and presented graphically in Figure 4.

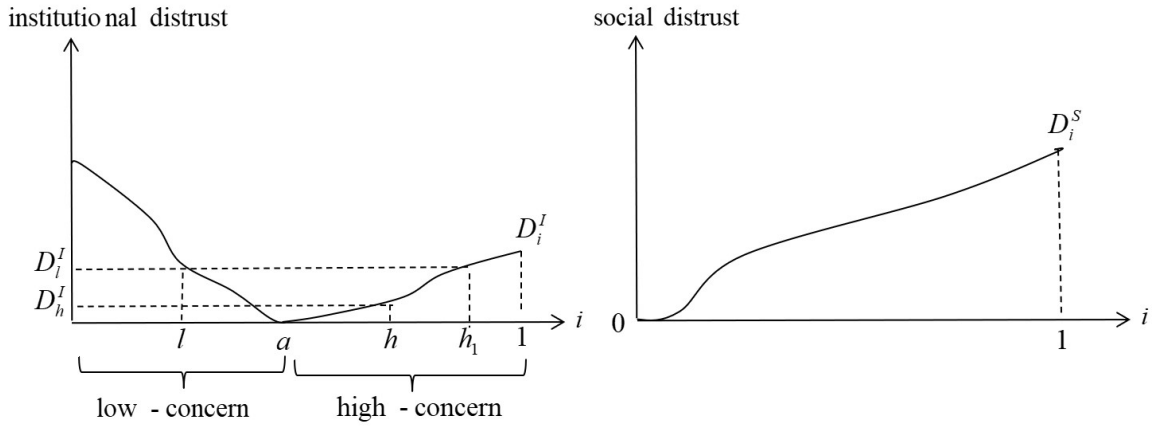
Proposition 1 (Asymmetry in trust). *Let $a \in [0, 1]$ be the individual whose entitled policy is equal to the government's policy, $\rho_a = \rho^*$, and let*

- $L \subset [0, 1]$ denote the subset of low-concern individuals such that $\rho_l < \rho^$, for any $l \in L$,*

²⁶Of course, moral hypocrisy may lead some non-complying individual to blame people that are complying more than herself. This is an interesting attitude to study but, as pointed out earlier, we omit it here.

- $H \subset [0, 1]$ denote the subset of high-concern individuals such that $\rho^* < \rho_h$, for any $h \in H$.
- i. For any ‘symmetric’ couple of low-concern/high-concern individuals, such that $|l - a| = |h - a|$, if the cost function is sufficiently convex and externality functions are sufficiently concave, the low-concern individual distrusts the government more than the high-concern individual: $D_l^I > D_h^I$;
- ii. The higher an individual’s level of concern, the more that individual distrusts other people: if $i > j$ then $D_i^S > D_j^S$.

Figure 4: Asymmetry in trust



Institutional distrust comes from disagreement with the government’s rule. So, if the government’s rule is unchanged, we expect overall institutional distrust to increase if opinions and concern become more heterogeneous across individuals. The more they disagree amongst themselves, the more they will blame the government because the distance between their most preferred rule and the government rule increases.

More heterogeneity in beliefs will lead to bigger differences in behavior, but also to more diversified expectations about what others ought to do. This in turn will lead to a bigger loss of social trust, which will come especially by the disappointment of the most concerned ones. Proposition 2 formalizes these results.

Proposition 2 (Disagreement). 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. TD^I and TD^S increase if

- i. heterogeneity in individuals’ levels of concern increases,
- ii. level of concern is sufficiently large and increases proportionally for all individuals.

Whereas proposition 2 focuses mainly on total levels of trust, the following two propositions explore how a given shock asymmetrically affects trust for subgroups being more or

less concerned. Those propositions will also guide our discussion of our empirical results in section 3. In this light, we are interested in two comparative statics: when the government rule is ‘perceived’ to increase, and when the costs of mitigation are ‘perceived’ to rise. Here we talk about perceptions because our experiment treatments in section 3 are aimed at increasing individuals’ perceptions about different aspects of the Covid crisis, such as the strictness of the rules or the cost of mitigating behavior.

Proposition 3 assumes the government rule increases. We do not model, for simplicity, why this should be the case. It might be due for instance to the fact that the government has got its own view about the severity of the externality and implements a paternalistic policy. If it becomes more concerned, the policy will be stricter. 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, an increase in the rule will lead all individuals to distrust others more, because with a stricter rule more individuals will not comply and take the risk of against being caught and fined, a result echoing Aghion et al. (2010).

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

- i. high-concerned individuals distrust institutions less and low-concerned distrust it more; the total effect on institutional distrust TD^I is positive if the perceived increase in ρ^* 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 studies how mitigation costs affect individual levels of trust. If costs are higher all individuals will think the rules should be less restrictive. Given the government rule, those who are very concerned will align more with the government and gain institutional trust, while the less concerned ones will disagree more and lose trust in the government. More individuals will choose not to comply or to further reduce their behavior, feeding the distrust of others.

Proposition 4 (Costly compliance). *If marginal mitigating costs $c'(b_i)$ are (perceived to be) increasing for any value of b_i , then, all things being equal, preferred mitigating policies ρ_i are lower for all individuals, so that*

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

2.5 Discussion

As we show in the following Section 3, Propositions 1 to 4 are consistent with our experimental evidence, but we think they can be given a general interpretation. Whenever a society is tackled by a new challenge or a crisis requiring social coordination, we might observe increasing distrust in institutions, especially by those who, for whatever reason, are less worried. Large disagreement, perhaps due to little knowledge about the severity of the crisis, might then lead to a large erosion of political trust. Additionally, our model predicts that more heterogeneity will lead to bigger differences in behavior. This in turn will further fuel social distrust in society. Here distrust comes especially from the most concerned ones, who are disappointed by the misbehavior of the less concerned ones.²⁷

Crises requiring stricter policies (Proposition 3) or stronger sacrifices (Proposition 4) are more likely to erode social trust but also risk undermining institutional trust. Also here, different groups might react differently to a given crisis or change in the policy, as the more concerned distrust others more, and the less concerned mostly lose trust in political institutions.

Take climate change, for example. We discussed above how beliefs about the severity of global warming and the effect of mitigating behavior can be very heterogeneous. Some individuals think 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 problem. 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. The latter kind of disappointment with the government comes especially from the low-concern groups that, according to our model, are expected to resist the implementation of those policies more. The stricter the rules are perceived to be, or the larger the costs of complying with them, the more these effects will play out.

All of these predictions are worrisome. 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. We will come back on this point in the conclusions.

²⁷If the government does very little, such as Bolsonaro or Trump at the outset of the Covid pandemic, our framework predicts that the concerned group will mostly develop distrust in institutions, while there will be little to complain about the non-virtuous behavior of the non-concerned ones.

3 Experimental Effects of the Covid-19 Crisis on Distrust

We now present a large online survey conducted in June 2020 and interpret its findings through the lens of our theoretical model.²⁸ The primary goal of this priming experiment was to make several dimensions of the Covid-19 crisis more salient to investigate its effect on individuals’ socio-political attitudes. 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 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 Taxation and the welfare state
Outcome questions Institutional trust Interpersonal trust Taxation and the welfare state	Treatment questions Sub-conditions: Health Health & Economic Health & Conflict
Further background information	Further background information

In line with other papers using similar priming techniques, such as [Alesina et al. \(2023\)](#) or [Bartoš et al. \(2021\)](#), our presumption is that the Covid-19 crisis becomes more salient by asking treatment questions before outcome questions. Any changes in treated respondents’ answers compared to the non-treated can then be attributed to thinking more intensely about the crisis, and proxy the way trust and political beliefs are affected.

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

1. **Health:** These questions tackle respondents’ day-to-day experience of the epidemic, e.g. which of the measures they have adopted, whether they had covid cases among

²⁸The 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 use for the multiple hypothesis testing in Appendix D.2.

The survey was pre-registered as AEARCTR-0005952. Heterogeneous treatment effects were not included in the pre-analysis plan. See Appendix E for a link to the questionnaire and a complete list of the outcomes, as well as Table D4 for summary statistics of all variables used.

their acquaintances and family members, whether they found the policy measures too strict or hard to bear, whether they were concerned for their health, etc. 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 whether they lost their job or any future job opportunities, whether people around them were suffering from economic distress, etc. These questions are expected to trigger respondents' awareness of the economic effects of the crisis, along with the (appropriateness of 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 more intensely about the social cohesion needed to overcome a crisis, and how important it is that people comply with the rules in order to do so.

The outcome questions elicit respondents' degree of institutional and interpersonal trust along with their attitudes towards redistributive policies:

Institutional trust: We ask respondents how much they trust politicians, the national government, and the European Union. In order to check for accuracy, we lastly introduce a costly task, similar to DellaVigna et al. (2016). We measure whether the respondents accept to spend time (5 minutes) to read a pro-European integration speech and give us 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 social trust using the standard question whether people can in general be trusted. We also measure support for direct democracy, using a question which elicits the degree to which respondents think 'the people, and not politicians, should make our most important policy decisions'. Whilst we initially included this question to size up populist attitudes, it can also capture a policy-oriented and hence indirect dimension of social trust, at least to some degree.³⁰

²⁹For a detailed description of this incentivised measure and its challenges see Appendix C.1 and C.3.

³⁰Indeed, respondents disagreeing with this statement will most likely think regular citizens are worse administrators than actual politicians, or believe both are equally bad. In both cases other citizens are not sufficiently trusted to delegate policy decisions to, which can be a proxy for lower trust in others. We will come back to this point later.

Taxation and the welfare state: We elicit respondents’ views on financing redistributive policies. In particular we measure approval of tax-financed poverty alleviation, levying taxes to provide public health care, to ensure adequate unemployment benefits and to provide a reasonable standard of living for the elderly. We also elicit perceived excessiveness of overall and own fiscal burden.

Importantly, we use the three treatment questions blocks to implement three mirroring treatments. 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&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&Cooperation’ questions could activate health-related aspects as well. To cleanly identify the impact of the Economic and the Conflict&Cooperation dimensions in the second and third treatments, we therefore intentionally activate and fix health aspects across all treatments.³¹

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,³² whilst κ denotes regional (NUTS-2) fixed effects.³³ 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 will pool all three treatments together, and refer to this as the "Covidfirst" condition. Summary statistics are reported in Table D4.

The survey was conducted in Germany, Italy, the Netherlands and Spain in the first two weeks of June 2020, in the aftermath of severe lockdown measures adopted in those four countries. A random sample of approximately 2,000 adult individuals per country were reached by the survey company *Respondi*. The sample was built to be close to representatives in each country (see Appendix C for further details about the survey flow and the data collection).

Appendix Table D18 shows that the randomization strategy is effective as there are no

³¹We 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&Cooperation conditions.

³²All 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 behavioural (incentivised) measure of support for the European Union (see Appendix C.3 for more details).

³³We 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.

sizeable differences in these variables between individuals in the Baseline groups and the Covidfirst groups. As further discussed in Appendix C.3, we address fatigue effects as well as demand effects by leaving the respondents blind to experimental conditions other than the one they are participating in. Appendix D.2 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.

3.1 Treatment Effects

We find that simply asking respondents questions about their pandemic experience before asking them about trust makes them less trustful towards others and political institutions. This is illustrated in Figure 5. 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, interpersonal trust drops -12% ($p < 0.01$) and support for direct democracy drops by -5% .³⁴ Primed respondents are 14% of a standard deviation less likely to read the Pro-EU speech. These results support the idea that when 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 D.³⁵

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

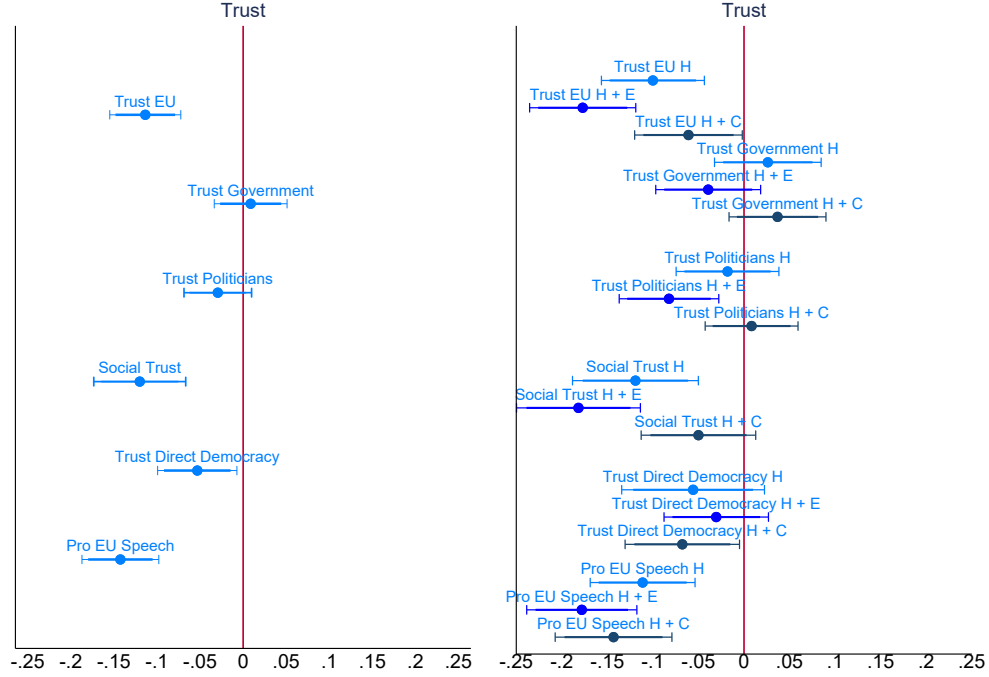
First, the treatment may have reminded respondents in each of the four countries that the government rules were quite strict during the first Covid wave in Spring 2020. This might have led concerned individuals to think that the government is right in setting such restrictive rules, while it might have 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 leading to a non-significant average effect that we see in the data. This explanation is consistent with 3-*i* saying that if the government rules become stricter, all other things being equal, concerned individuals will trust the government more while unconcerned individuals will trust the government less.

As for the second explanation, recall that the left panel pools the three sub-conditions together. So it might be that the three sub-conditions lead respondents to focus on different

³⁴As discussed above, such a negative effect on support for direct democracy can be interpreted as an indirect drop of social trust along the policy-dimension. The fact that it significantly coincides with the overall drop in social trust underpins this interpretation empirically. The null-effect we find for the other questions in the larger survey measuring populist attitudes speaks in the same direction (see Appendix E).

³⁵We 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.

Figure 5: 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).

dimensions of the Covid crisis, with offsetting effects on trust in government/politicians. The right panel of Figure 5 checks for this possibility. It breaks the treatment effect down into the three sub-conditions. Trust in government and in politicians decreases (significantly) in the ‘Economic’ sub-condition (H+E) while it does not in the ‘Conflict and Cooperation’ sub-condition (H+C) and in the ‘Health’ sub-condition (H). In fact, the non-significant effect in the left panel is the composition of three opposite effects displayed in the right panel. It is plausible that the ‘Economic’ sub-condition (H+E) has led respondents to focus on the sizeable economic costs associated to the crisis. For instance, treated individuals might think that job losses are due to costly lockdown measures more so than non-treated ones. This would be consistent with Proposition 4-*i* saying that higher mitigating costs lead all individuals to think that mitigating measures should be less severe. If this is the case, unconcerned individuals trust the government less, while concerned individuals trust the government more. The overall effect on trust in governments and politicians is negative if the priming effect on costs is sufficiently large, as we see in the H+E sub-condition.

Interestingly, the negative effect on social trust is stronger in the ‘Economic’ sub-condition, compared to the other two treatments. This is consistent with the idea that when a respondent thinks more intensely about economic hardship she might blame others more because their non-compliant behavior causes more social harm. This result is then consistent with Proposition 4-*ii*.

Inversely, the ‘Conflict and Cooperation’ sub-condition might have led respondents to think more intensely that the crisis is a common threat and it is important to have a strong and trustworthy government setting common rules. In this sense, the treatment might have reduced the heterogeneity of respondents’ views about the right policy to implement. As suggested by Proposition 2-*i*, lower heterogeneity yields less political distrust, a prediction which is in line with H+C effect in the right panel of Figure 5. Similarly, we also see that H+C effect on social trust is smaller than for the other two treatments.

We will unpack these explanations further in our heterogeneity analysis in the next subsection.

3.2 Heterogeneity Effects

We first of all investigate the heterogeneous treatment effects along the dimension of individual concern with the crisis. As a part of the ‘Health’ treatment, individuals are asked how much they are worried about their health due to the epidemic (on a scale from 1 to 10). We use answers to study the interaction between health-concern and the treatment. Table 2 again pools the three treatment sub-conditions.

As for social trust, the interaction term in the third row shows that the treatment effect is stronger for more concerned individuals (columns 4-5). The more an individual is concerned, the more the treatment results in a loss of social trust. This result is consistent with the theoretical predictions in Propositions 3-*ii* and 4-*ii*, and with our interpretation of trust in direct democracy as a proxy for social trust as explained above.

Table 2: Priming Covid, Trust and Health Concerns

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Trust Dir. Democracy	(6) Pro EU Speech
Covidfirst	-0.116*** (0.020)	0.007 (0.021)	-0.031 (0.020)	-0.119*** (0.027)	-0.054** (0.023)	-0.143*** (0.022)
Health Concerns	0.109*** (0.023)	0.067*** (0.022)	0.068*** (0.022)	-0.024 (0.022)	0.071*** (0.025)	0.032** (0.015)
Covidfirst * Health Concerns	0.007 (0.029)	0.058** (0.025)	0.060*** (0.023)	-0.041* (0.024)	-0.069*** (0.026)	0.037* (0.019)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.058	0.156	0.214	0.079	0.061	0.087
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. ‘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, equalised 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

As for trust in government and politicians, the first row shows no significant effect, in line with the left panel of Figure 5. However, the third row reports a significant positive effect for highly-concerned individuals, as is the case for the incentivised outcome measuring

(EU) political support. This is consistent with the possibility that the treatments have fed respondents' perception that the government rules were actually very strict or that the mitigating costs were quite high. So it corroborates the two explanations in the previous subsection and it is in line with Propositions.3-*i* and 4-*i*. There is no heterogeneity effect for Trust in the EU specifically.

In Table 3, we replicate this analysis using a discrete measure of health concern, by splitting this variable in three tertiles. It is immediately clear that the negative treatment effects on trust in institutions are driven by very unconcerned individuals (first tertile), while the negative effects on interpersonal trust are driven by very concerned individuals (top tertile). This further corroborates the theory, along with the idea that the treatments effects travel through the perception that rules were strict and costs were high.

Table 3: Priming Covid, Trust and Health Concerns: tertiles

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Trust Dir. Democracy	(6) Pro EU Speech
Covidfirst	-0.155*** (0.038)	-0.071* (0.037)	-0.111*** (0.033)	-0.069 (0.042)	0.003 (0.039)	-0.208*** (0.039)
Covidfirst * Health Concerns 2 nd Tertile	0.105** (0.050)	0.122** (0.052)	0.128** (0.052)	-0.058 (0.058)	-0.063 (0.054)	0.109** (0.054)
Covidfirst * Health Concerns 3 rd Tertile	0.012 (0.068)	0.125** (0.063)	0.124** (0.059)	-0.108* (0.056)	-0.124* (0.064)	0.099** (0.047)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.056	0.156	0.213	0.080	0.062	0.089
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. '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, equalised 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 D11 in Appendix reports similar findings when heterogeneity is captured by self-reported compliance with rules, measured by another 'Health' treatment question. We find that respondents reporting higher compliance lose more trust in others and trust the government and politicians more. This is not surprising since the theory predicts that people comply more if they are more concerned, and Table 4 indeed shows a strong correlation between compliance and health-concern (0.42).

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 their answers as respondents' expectations about the civic behavior of others. We call this variable 'Civic Expectations', and find that it is highly correlated

with health-concern (0.37) and with self-reported compliance (0.48).³⁶ Those expecting others to behave in a civic manner are likely to report health concern and self-compliance with the rules. Somehow importantly, if we use these expectations to capture heterogeneity of concern, we find that the treatment yields a sharp drop in social trust amongst those expecting more civic behavior. This result supports our basic theoretical idea that social distrust is caused by disappointed expectations about the civic behavior of others (cf. Table D12 in Appendix).

Table 4 sheds further light on the mechanism connecting individuals' concern, compliance, and civic-expectations with their perceptions of the government's management of the pandemic. Our survey included two specific 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' behavior. This validates our theoretical channels predicting 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 civic behavior from others.

Table 4: Evaluations of Covid Regulations Strictness

	(1)	(2)	(3)	(4)	(5)	(6)
	Rules too Strict	Rules too Strict	Rules too Strict	Gov. Performance	Gov. Performance	Gov. Performance
Compliance	-0.703*** (0.043)			0.192*** (0.015)		
Health Concerns		-0.261*** (0.046)			0.116*** (0.016)	
Civic Expectations			-1.201*** (0.059)			0.335*** (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 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 regulations. 'Civic Expectations': factor variable increasing in expectations of civic behaviour of others. 'Health Concerns': self-perceived Covid 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 5 investigates the heterogeneous effects of health-concern along the three sub-conditions. We can use these results to further underpin our interpretation of the experimental outcomes in Section 3.1. We advanced the interpretation that the Economic sub-

³⁶The 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 Health & Conflict receiving this set of questions.

condition (H+E) would have led individuals to think intensely about the huge costs of the pandemic and the restrictiveness of the government rules. This would have led unconcerned individuals to trust the government less, and concerned individuals to trust the government more. This interpretation is supported by the effect of interaction between concern and the H+E sub-condition (the sixth row, columns 2 and 3), showing that in the Economic treatment more concerned respondents gain trust in the government once primed. Thus, the significant negative effect of H+E on political trust we found in the right panel of Figure 5 is driven by the loss of trust experienced by unconcerned individuals, as predicted by Proposition 4-*i*. in the case of a large increase in adjustment costs.

Moreover, Table 5 shows that also the 'Conflict and Cooperation' sub-condition (H+C) yields a significant and positive effect on the institutional trust of concerned respondents (seventh row, columns 2 and 3). Thus the high-concerned turn to the government more than the low-concerned, once they are reminded of the importance of cooperation. The same respondents also trust direct democracy less (Column 5), which indicates they distrust other people more when thinking of them as directly taking part in political decision-making.³⁷

Table 5: Priming Covid, Trust, Health Concerns: by treatment

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Trust Dir. Democracy	(6) Pro EU Speech
T = 1 (H)	-0.102*** (0.028)	0.028 (0.030)	-0.018 (0.028)	-0.120*** (0.035)	-0.058 (0.040)	-0.108*** (0.029)
T = 2 (H+E)	-0.185*** (0.028)	-0.045 (0.029)	-0.086*** (0.027)	-0.185*** (0.034)	-0.032 (0.029)	-0.182*** (0.029)
T = 3 (H+C)	-0.060** (0.029)	0.037 (0.026)	0.010 (0.025)	-0.052* (0.032)	-0.071** (0.031)	-0.140*** (0.031)
Health Concerns	0.109*** (0.023)	0.067*** (0.022)	0.069*** (0.022)	-0.024 (0.022)	0.070*** (0.025)	0.032** (0.015)
T = 1 * Health Concerns	-0.022 (0.033)	0.024 (0.030)	0.022 (0.028)	-0.058 (0.036)	-0.026 (0.036)	0.024 (0.025)
T = 2 * Health Concerns	0.023 (0.035)	0.066* (0.034)	0.072** (0.032)	-0.048 (0.030)	-0.107*** (0.033)	0.029 (0.025)
T = 3 * Health Concerns	0.020 (0.040)	0.085** (0.035)	0.086*** (0.030)	-0.015 (0.031)	-0.072** (0.030)	0.059** (0.027)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.060	0.157	0.215	0.081	0.062	0.088
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). '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

³⁷The effect for social trust suffers from the loss in power by splitting the sample in three to focus on sub-treatments. Nevertheless, the H+E and H sub-conditions mark a close to significant effect (p-values of 0.1 and 0.11)

3.3 Taxation & Welfare State

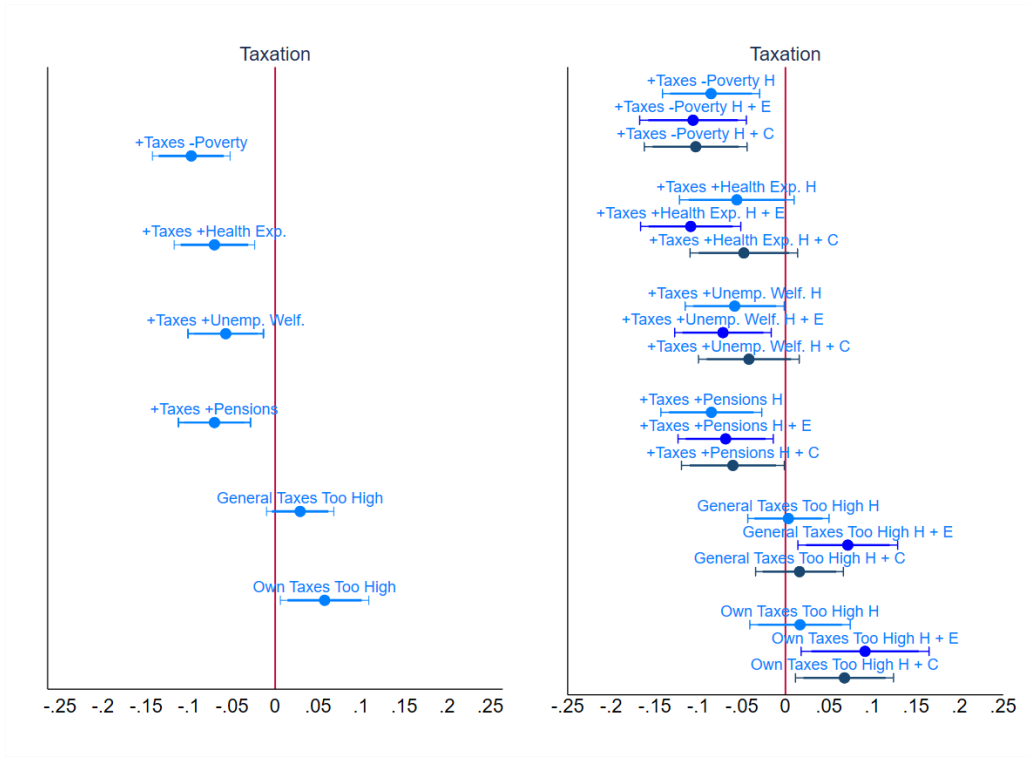
As briefly discussed in the introduction, a large literature in social science studies the importance of institutional and social trust for the well-being of modern democracies. One channel through which trust affects social progress is the willingness of trustful people to contribute to the welfare state. In this paper, we have argued that crises erode institutional trust, as policies are seen as intrusive or ineffective. We have not tackled the relationship between institutional trust and preferences for redistributive taxation theoretically. However, it seems plausible to hypothesize that if trust in institutions is lower, people are less willing to contribute to the welfare-state bill. Crises such as Covid-19 can then translate into a decreased willingness to pay, as the political class running welfare programs are trusted less (Rothstein, 1998). A similar argument is plausible if social trust declines, as it buttresses the belief that others are less worthy of the welfare system benefits (Daniele and Geys, 2015). We can use our survey experiment to explore the relationship between awareness of Covid-19 crisis and preferences over redistribution.

The left panel of Figure 6 (Table D7 in appendix) shows the treatment effect (the three sub-conditions together) on several dimensions of preferences over redistributive policies. Thinking intensely about Covid-19 lowers people’s agreement with the principle of using taxation to alleviate poverty, to provide social insurance and to finance health-care and pensions. Treated respondents are more likely to think that the tax bill is too high. From the right panel (Table D8 in appendix), we see that these effects are driven mostly by the ‘Economic’ sub-condition (H+E). As Table 6 shows, these results hold both for concerned and non-concerned respondents. However, our analysis in Section 3.2. suggests that the mechanisms behind this decline differ between the two groups. In Table 2 we show that our treatments lead to a decline in social trust predominantly among concerned respondents, while institutional trust declines among the non-concerned. Since, as we explain in the above paragraph, both institutional and social trust matter for welfare state support, this measure goes down among all respondents, irrespective of their degree of concern about the virus.

However, effects are relatively strong also in the ‘Conflict and Cooperation’ (H+C) sub-condition. Recall that the H+C treatment has no significant negative effect on trust in government and politicians as well as social trust. This means that negative treatment effects on taxation travel through additional channels than just lack of trust. It therefore seems plausible that all sub-conditions lead people to believe that the crisis will imply a massive increase in public spending. Once they focus on the crisis, they do not want higher taxation as a result. This might lead citizens to support large public debt expansion programs as a way to escape increases in current taxation, as also found by Rees-Jones et al. (2022) for the United States specifically. Our results then also offer a political-economy explanation of why Western governments extensively engaged in such programs in the aftermath of the

Covid-19 crisis.

Figure 6: Treatment effects on taxation/welfare state



Left panel: all sub-conditions pooled 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).

Table 6: Priming Covid, Tax and Health Concerns

	(1)	(2)	(3)	(4)	(5)	(6)
	Tax Poverty	Tax Health Exp.	Tax Unemp. Welf.	Tax Pensions	General Tax	Own Tax
Covidfirst	-0.102*** (0.023)	-0.074*** (0.024)	-0.061*** (0.023)	-0.074*** (0.021)	0.033 (0.020)	0.058** (0.026)
Health Concerns	0.094*** (0.023)	0.111*** (0.021)	0.105*** (0.024)	0.110*** (0.021)	0.050** (0.021)	0.095*** (0.020)
Covidfirst*Health Concerns	0.047* (0.026)	0.011 (0.023)	0.045* (0.026)	0.028 (0.024)	-0.037* (0.021)	-0.056** (0.022)
Observations	8,234	8,234	8,234	8,234	8,234	8,234
R-squared	0.043	0.070	0.080	0.067	0.135	0.121
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). 'Covidfirst': all sub-conditions 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, equalised 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

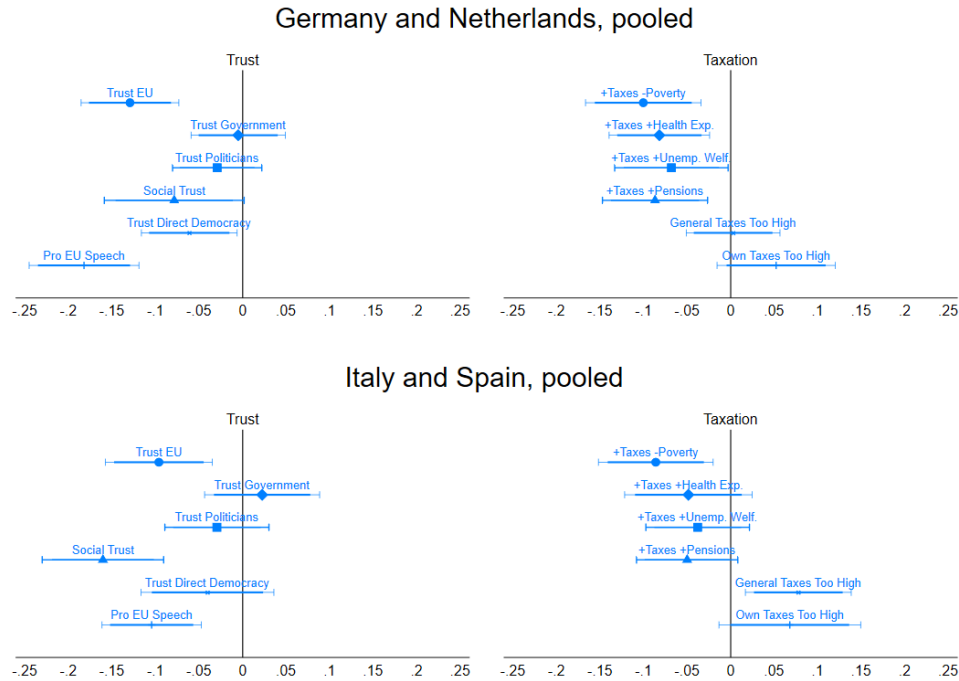
3.4 Country Level Analysis

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 couples 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.³⁸ This is not surprising, as Proposition 2 predicts that higher overall levels of concern bring about higher levels of institutional and interpersonal distrust.

Given these differences in initial conditions, we decided to replicate our analysis splitting the sample in two across these country lines in Figure 7, and in Tables D9 and D10 in Appendix. What immediately stands out in the left panels of Figure 7 is that the treatment effect on social trust is more pronounced in Italy and Spain than in Germany and the Netherlands. This is again consistent with our theory. Following Proposition 3, the treatment might have led all respondents to think of rule stringency, with stronger negative effects on the social trust of the more concerned. This is then expected to lead to larger drops in Spain and Italy, where people were more worried, than in Germany and Netherlands.

Figure 7: Country Level Analysis



Note: The figures display the impact of Covidfirst on our set of outcomes concerning trust and taxation preferences 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.

³⁸The 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.

In Table D15 lastly, we replicate the heterogeneity analysis for the same country subsets. The interaction effects for social and institutional trust are similar across the two groups.

3.5 Placebo Test

As mentioned above, the original survey included a wider set of outcomes which were not related to the theoretical model. These are questions on other dimensions of trust (police and media), identity (to own town, country and European Union), policy preferences related to populism (immigration, free media, having a strong leader in power, and excessive elite power), giving up freedom in exchange for more safety, and values related to globalization and local traditions. In this case, the model does not predict any heterogeneous effects depending on the level of Covid-related health concerns: we therefore use them as Placebo outcomes. In Figure D6 we replicate the model of Table 2 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.

4 Concluding Remarks

In this paper we argued that distrust is fueled by the disappointment of individuals' expectations about what government should do and how others should behave. It then increases when expectations diversify or polarize.

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.

We argued that in an increasingly complex world, interdependence increases and so does the demand for new rules to coordinate individual behavior. At the same time, rules become less simple and individuals' interests are more diversified, with the result that individuals will have more diversified expectations of what government and citizens should do. So complexity, interdependence, and polarized political preferences bring about the risk of trust decline.

We studied externalities, regulation, and distrust. We think however that our framework can be fruitfully extended to explore the link between taxation and distrust. Here the natural heterogeneity across individuals' preference for taxation and welfare state would be income levels, with the poorest people expected to welcome any increase in redistributive taxation. However, reality shows that even the poor tend to distrust the government more when it raises taxes. Indeed, our survey shows that all respondents agreed on reducing the tax burden once primed with Covid, a result suggesting that trust crises lead both rich and poor to demand lower taxes, more public debt, and defer expensive political measures.

We assumed that individuals' opinions are exogenous and directly map onto their expectations. This 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, 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.

We realistically assumed that reduced compliance with government rules brings about interpersonal distrust. This relation can however go the opposite way too, giving rise to another endogenous mechanism by which distrust creates the conditions for new conflicts and further the loss of institutional and interpersonal trust.

The ideas developed in this paper can contribute to a more general theory of political participation that applies to voting, protests, and other political activities. Voters, for example, can be more easily mobilized towards protest voting when they distrust institutions. Political leaders, by exploiting the emotional reactions of disappointed voters, can more easily 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 or reduce its prerogatives. The relationship between distrust, partisan identification, and polarization of political preferences is important and requires further investigation.

Our model suggests that information manipulation plays a key role in generating distrust. More information is not necessarily a good thing. 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, a powerful channel through which the media influence the political dynamic.

What to do then? Our paper does not take a normative stance. It focuses on the causes of distrust. The ideas developed here are, however, useful to shape trust-building strategies.

We focused on the 'dark side' of too much heterogeneity of political opinions. A strategy to rebuild trust should then aim at reducing non-dialectical opposition of political opinions. Of course, that is a very easy task to say but terribly difficult to achieve. Sociologists and communication experts have underlined the rules of social dialogue. The narrative should be less divisive, and aimed to build consensus and tolerance rather than conflict. Citizens should be educated to discriminating between information with scientific basis and to being suspicious of all others. At the same time, journalists and social media should check information sources and avoid language that is divisive or fuels hatred or anger. Economic

research can contribute a lot to identifying how these mechanisms work and when they are more effective. Preventing trust erosion is vital for the future of 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.
- Aksoy, C. G., Eichengreen, B., Saka, O., et al. (2020). The political scar of epidemics. Technical report, Institute of Labor Economics (IZA).
- Alesina, A., Baqir, R., and Hoxby, C. (2004). Political jurisdictions in heterogeneous communities. *Journal of political Economy*, 112(2):348–396.
- Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, 43(3):762–800.
- Alesina, A. and La Ferrara, E. (2000). Participation in heterogeneous communities. *The quarterly journal of economics*, 115(3):847–904.
- Alesina, A., Miano, A., and Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, 90(1):1–39.
- Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.
- Alesina, A., Stantcheva, S., and Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, 108(2):521–54.
- 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., Beasley, E., Cohen, D., and Foucault, M. (2018). The Rise of Populism and the Collapse of the Left-Right Paradigm: Lessons from the 2017 French Presidential Election.
- Algan, Y. and Cahuc, P. (2009). Civic virtue and labor market institutions. *American Economic Journal: Macroeconomics*, 1(1):111–45.
- Algan, Y., Guriev, S., Papaioannou, E., and Passari, E. (2017). The european trust crisis and the rise of populism. *Brookings Papers on Economic Activity*, 2017(2):309–400.
- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., and Yang, D. (2020a). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of public economics*, 191:104254.
- Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020b). The welfare effects of social media. *American Economic Review*, 110(3):629–76.
- Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

- Ashraf, Q. and Galor, O. (2013). Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, 103(3):528–33.
- Barsbai, T., Licuanan, V., Steinmayr, A., Tiongson, E., and Yang, D. (2020). Information and the acquisition of social network connections. Technical report, National Bureau of Economic Research.
- Bartoš, V., Bauer, M., Cahlíková, J., and Chytilová, J. (2021). Covid-19 crisis and hostility against foreigners. *European Economic Review*, 137:103818.
- Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.
- Besley, T. J. and Persson, T. (2022). The political economics of green transitions.
- Boeri, M. T., Mishra, M. P., Papageorgiou, M. C., and Spilimbergo, M. A. (2018). *Populism and Civil Society*. International Monetary Fund. Google-Books-ID: EKAZEAAAQBAJ.
- Bonomi, G., Gennaioli, N., and Tabellini, G. (2021). Identity, beliefs, and political conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411.
- Bor, A., Jørgensen, F., and Petersen, M. B. (2022). Discriminatory attitudes against the unvaccinated during a global pandemic. *Nature*, pages 1–3.
- Bordalo, P., Tabellini, M., and Yang, D. Y. (2020). Issue salience and political stereotypes. Technical report, National Bureau of Economic Research.
- Brodeur, A., Gray, D. M., Islam, A., and Bhuiyan, S. (2020). A literature review of the economics of covid-19.
- Bursztyn, L., Egorov, G., and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American economic review*, 110(11):3522–3548.
- Bursztyn, L., Rao, A., Roth, C., and Yanagizawa-Drott, D. (2022). Opinions as Facts. *The Review of Economic Studies*.
- Canen, N., Kendall, C., and Trebbi, F. (2020). Unbundling polarization. *Econometrica*, 88(3):1197–1233.
- Daniele, G., Aassve, A., and Moglie, M. L. (2023). Never forget the first time: The persistent effects of corruption and the rise of populism in italy. *Journal of Politics*.
- Daniele, G. and Geys, B. (2015). Interpersonal trust and welfare state support. *European Journal of Political Economy*, 39:1–12.
- de Quidt, J., Johannes, H., and Christopher, R. (2018). Measuring and Bounding Experimenter Demand. *American Economic Review*, 108(11):64.

- Delhey, J., Dragolov, G., Boehnke, K., et al. (2018). Social cohesion and well-being in europe.
- DellaVigna, S. and Gentzkow, M. (2010). Persuasion: empirical evidence. *Annu. Rev. Econ.*, 2(1):643–669.
- DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.
- DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1):143–181.
- Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2017). Culture, ethnicity, and diversity. *American Economic Review*, 107(9):2479–2513.
- Di Tella, R., Dubra, J., and Lagomarsino, A. (2021). Meet the oligarchs: Business legitimacy and taxation at the top. *The Journal of Law and Economics*, 64(4):651–674.
- Durante, R., Guiso, L., and Gulino, G. (2021). Asocial capital: Civic culture and social distancing during covid-19. *Journal of Public Economics*, 194:104342.
- Dustmann, C., Eichengreen, B., Otten, S., Sapir, A., Tabellini, G., and Zoega, G. (2017). Europe’s trust deficit. *Causes and Remedies. London: Centre for Economic Policy Research*.
- Easterly, W., Alesina, A. F., and Baqir, R. (1997). Public goods and ethnic divisions. *Available at SSRN 46092*.
- Enke, B. (2019). Kinship, cooperation, and the evolution of moral systems. *The Quarterly Journal of Economics*, 134(2):953–1019.
- Enke, B. (2020). Moral values and voting. *Journal of Political Economy*, 128(10):3679–3729.
- Foremny, D., Sorribas-Navarro, P., and Vall Castelló, J. (2020). Living at the peak: Health and public finance during the covid-19 pandemic. *Available at SSRN 3578483*.
- Friedman, B. M. (2005). *The Moral Consequences of Economic Growth*. Alfred A. Knopf, New York.
- Fukuyama, F. (2018). *Identity: The demand for dignity and the politics of resentment*. Farrar, Straus and Giroux.
- Genschel, P., Hemerijck, A., Russo, L., and Nasr, M. (2021). Solidarity and trust in times of covid-19. *Robert Schuman Centre for Advanced Studies Research Paper No. RSC*.
- 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).
- Giuliano, P. and Wacziarg, R. (2020). Who voted for trump? populism and social capital. Technical report, National Bureau of Economic Research.
- Glaeser, E. L., Ponzetto, G. A., and Shleifer, A. (2007). Why does democracy need education? *Journal of economic growth*, 12(2):77–99.
- Guiso, L., Herrera, H., Morelli, M., and Sonno, T. (2020). Economic insecurity and the demand of populism in europe. *Einaudi Institute for Economics and Finance*.
- Guiso, L., Sapienza, P., and Zingales, L. (2004). The role of social capital in financial development. *American economic review*, 94(3):526–556.
- Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic perspectives*, 20(2):23–48.
- Guriev, S. (2018). Economic drivers of populism. In *AEA Papers and Proceedings*, volume 108, pages 200–203.
- Guriev, S. and Papaioannou, E. (2022). The political economy of populism. *Journal of Economic Literature*, 60(3):753–832.
- Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large italian firm. *The Quarterly Journal of Economics*, 115(3):1057–1090.
- Knack, S. and Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & politics*, 7(3):207–227.
- Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly journal of economics*, 112(4):1251–1288.
- Kuziemko, I., Norton, M. I., Saez, E., and Stantcheva, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508.
- La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (1999). The quality of government. *The Journal of Law, Economics, and Organization*, 15(1):222–279.
- List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.

- Maggi, G. and Staiger, R. W. (2022). International climate agreements and the scream of greta. Technical report, National Bureau of Economic Research.
- Martinangeli, A. F. M., Povitkina, M., Jagers, S., and Rothstein, B. (2023). Institutional Quality Causes Generalized Trust: Experimental Evidence on Trusting under the Shadow of Doubt. *American Journal of Political Science*. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12780>.
- Norris, P. and Inglehart, R. (2019). *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press.
- 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.
- Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography*, 6(1):9–44.
- 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. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian political studies*, 30(2):137–174.
- Putnam, R. D. et al. (2000). *Bowling alone: The collapse and revival of American community*. Simon and schuster.
- Ramos-Toro, D. (2019). Interpersonal diversity and carbon emissions. *Mimeographed*.
- Rees-Jones, A., D’Attoma, J., Piolatto, A., and Salvadori, L. (2022). Experience of the covid-19 pandemic and support for safety-net expansion. *Journal of economic behavior & organization*, 200:1090–1104.
- Rothstein, B. (1998). *Just institutions matter: The moral and political logic of the universal welfare state*. Cambridge university press.
- Sarracino, F. (2012). Money, sociability and happiness: are developed countries doomed to social erosion and unhappiness? *Social Indicators Research*, 109(2):135–188.

- Stevenson, B. and Wolfers, J. (2011). Trust in public institutions over the business cycle. *American Economic Review*, 101(3):281–87.
- Tabellini, G. (2008a). Institutions and culture. *Journal of the European Economic association*, 6(2-3):255–294.
- Tabellini, G. (2008b). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Appendices

A The Specific Case of a Pandemic Crisis

To clarify, and to illustrate our empirical application further, we bring the general model to the context of a pandemic crisis and simplify the model by introducing linear benefit functions and quadratic costs. The population size is still normalised to unity, and individuals' types are uniformly distributed in $[0, 1]$. The virtuous dimension of individual behaviour $b_i \in [0, \infty)$ can in this specific case be thought of as mask wearing, respecting lockdown, quarantine or social distancing measures, getting a vaccine, etc. Doing so, individuals protect both themselves and others, so that b_i again has private and public benefits. Atomistic individuals ignore however the positive effect of their behaviour on others. In this parametric version we assume the benefit function of each individual i is linear,

$$\varepsilon_i(b_i + B_{-i}) = i \cdot (b_i + B_{-i}),$$

where $i > 0$ is both each individual's identity and her level of concern. The private cost of protective behaviour is now expressed by $c(b_i) = \frac{1}{2}(b_i)^2$, so individual utility becomes

$$U_i = i(b_i + B_{-i}) - \frac{1}{2}(b_i)^2. \quad (22)$$

Optimal decentralised behaviour, b_i^* , solves the FOC, $i - b_i = 0$, which yields an interior solution as the SOC is $-1 < 0$. Note how linear benefit functions greatly simplify the analysis, since decentralized behavior does not depend on the behavior of others. It yields no strategic substitutability. All individuals have their own decentralized dominant strategy, $b_i^* = i$. Still each individual i does not take into account her positive contribution to the overall externality. Given $B_{-i}^* = \int_0^1 j dj = \frac{1}{2}$, individual utility in a decentralized equilibrium thus becomes

$$U_i = i \left(i + \frac{1}{2} \right) - \frac{1}{2}i^2,$$

which leaves scope for policy intervention.

A.1 Entitled Policies

Individuals have their own entitled policy, ρ_i^* , maximizing their indirect utility function $V_i(\rho)$. If $\rho < 1$, some individuals will need to increase their behavior up to the rule while

some others will not. Thus, if individual i is an affected individual,³⁹

$$V_i(\rho) = i \left(\rho + \int_0^\rho \rho dj + \int_\rho^1 j dj \right) - \frac{1}{2}(\rho)^2,$$

The FOC is $h_i + \rho h_i - \rho = 0$ and the SOC is negative. So $\rho_i^* = \frac{i}{1-i}$ which is lower than one if $i < 1/2$. If instead $\rho \geq 1$ all individuals will have to comply, so the objective function is

$$V_i(\rho) = i \left(\rho + \int_0^1 \rho dj \right) - \frac{1}{2}(\rho)^2,$$

which is maximized by $\rho_i = 2i$. Summing up,

$$\begin{cases} \rho_i^* = \frac{i}{1-i} & \text{if } i < 1/2 \\ \rho_i^* = 2i & \text{if } i \geq 1/2 \end{cases}$$

Note that $\frac{\partial \rho_i^*}{\partial i} > 0$: more concerned individuals want stricter policy.

A.2 The Government Policy

The government maximizes a utilitarian welfare function in two steps. First, it assumes citizens fully obey by the rule. Second, it enforces the policy through a fine and monitoring activity.

First step: the objective function is

$$W(\rho) = \begin{cases} \int_0^\rho i \left(\rho + \int_0^\rho \rho dj + \int_\rho^1 j dj \right) - \frac{1}{2}(\rho)^2 di + \int_\rho^1 i \left(i + \int_0^\rho \rho dj + \int_\rho^1 j dj \right) - \frac{1}{2}i^2 di & \text{if } \rho < 1 \\ \int_0^1 i (\rho + \rho) - \frac{1}{2}(\rho)^2 di & \text{if } \rho \geq 1 \end{cases}$$

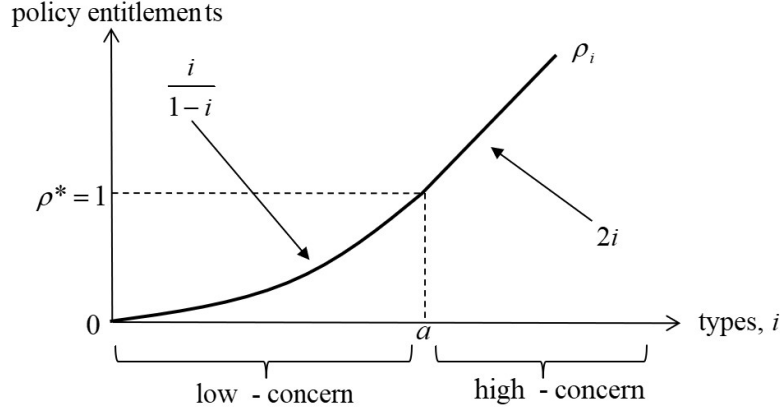
If $\rho < 1$ the welfare function is always increasing, so it is maximized in the corner, $\rho^* = 1$, while if $\rho \geq 1$ the stationary point is $\rho^* = 1$. So the socially optimal policy is unique and is $\rho^* = 1$. Note that this is also the most preferred rule by the average-concern individual $a = \frac{1}{2}$. Low-concerned individuals with $i < \frac{1}{2}$ want a more lenient rule, while the high-concern ones want a stricter rule (See Figure 5).

The government chooses monitoring activity, m^* , taking into account citizens' reaction. By (15), individual i chooses her behavior $b_i^*(\rho^*, m)$ so as to maximize her utility function,

$$U_i(b_i, \rho^*, m) = i(b_i + B_{-i}) - \frac{1}{2}b_i^2 - \bar{k} \cdot m \frac{(1 - b_i)^+}{1}$$

³⁹In fact in this linear model i is always affected; i.e., she always chooses a rule that is higher than her type. To see this, suppose i was unaffected (i.e., $\rho_i^* < i$). Her utility function would be $V_i(\rho) = i \left(i + \int_0^\rho \rho dj + \int_\rho^1 j dj \right) - \frac{1}{2}i^2$. It is increasing in ρ , thus $\rho_i^* > i$, which is a contradiction implying that i cannot be unaffected.

Figure A1: Overview of the linear case

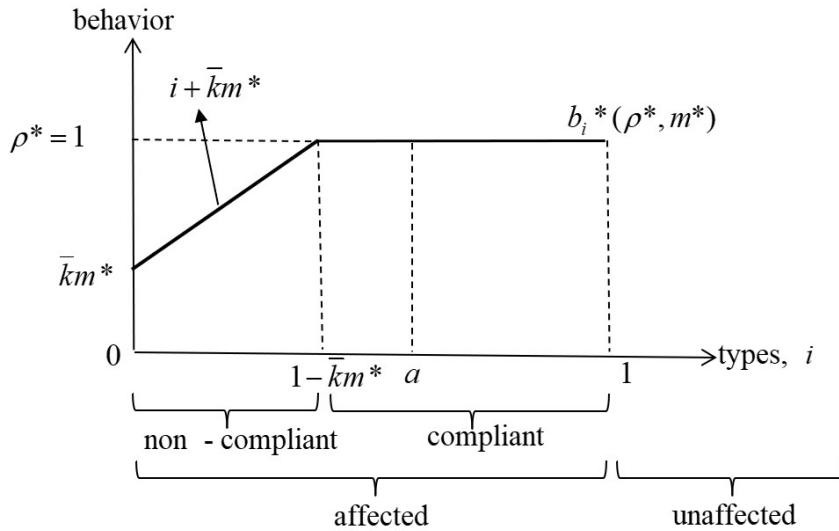


Optimal behavior is then

$$b_i^*(1, m) = \begin{cases} i + \bar{k}m & \text{if } i < 1 - \bar{k}m \\ 1 & \text{if } i \geq 1 - \bar{k}m \end{cases} \quad (23)$$

The government chooses optimal monitoring, m^* , taking into account the above individual reaction. We will skip this step for simplicity. We just assume that \bar{k} and m^* are such that $\bar{k}m^* < 1$. By (23), population is split in two groups. Those choosing to fully comply, $b_i^*(1, m^*) = \rho^* = 1$, in order not to pay the fine, and those choosing not to fully comply, $b_i^*(1, m^*) = 1 - \bar{k}m^* < \rho^* = 1$. Differently from the general model, here we have only two groups, because we constraint concern levels in $[0, 1]$ (See Figure 5).

Figure A2: Equilibrium behaviour linear case



A.3 Institutional and Social Distrust

By (20) institutional distrust is the following

$$D_i^I = \begin{cases} \left[i \left(2i + \int_0^1 2idj \right) - \frac{1}{2}(2i)^2 \right] - \left[i \left(1 + \int_0^1 dj \right) - \frac{1}{2} \right] = 2i^2 - 2i + \frac{1}{2} & \text{if } i \geq \frac{1}{2} \rightarrow \rho_i^* \geq \rho^* \\ \left[i \left(\frac{i}{1-i} + \int_0^{\frac{i}{1-i}} \frac{i}{1-i} dj + \int_{\frac{i}{1-i}}^1 j dj \right) - \frac{1}{2} \left(\frac{i}{1-i} \right)^2 \right] - \left[i \left(1 + \int_0^1 dj \right) - \frac{1}{2} \right] \\ = \frac{i^2}{1-i} + \frac{i^3}{(1-i)^2} + \frac{i-2i^2}{2(1-i)^2} - \frac{i^2}{2(1-i)^2} - 2i + \frac{1}{2} & \text{if } i < \frac{1}{2} \rightarrow \rho_i^* < \rho^* \end{cases}$$

The first line refers to a high-concern individual $i > \frac{1}{2}$ whose entitled policy is $\rho_i^* = 2i$. She thinks the government rule is too lenient. The second line concerns a low-concern individual $i < \frac{1}{2}$, with $\rho_i^* = \frac{i}{1-i}$. This individual thinks the government rule is too strict. It can be verified that distrust is increasing in i for high-concern individuals, while it is decreasing for low-concern individuals $i < \frac{1}{2}$. As predicted by Proposition 1.i., any low-concerned individual distrusts government more compared to a “symmetric” high-concern one.

By (21) social distrust is the following

$$D_i^S = \begin{cases} \left[i \left(1 + \int_0^1 dj \right) - \frac{1}{2}i^2 \right] - \left[i \left(1 + \int_0^{1-\bar{k}m^*} j + \bar{k}m^* dj + \int_{1-\bar{k}m^*}^1 dj \right) - \frac{1}{2}i^2 \right] = i - 2\bar{k}m^*i + (\bar{k}m^*)^2 i & \text{if } i \geq 1 - \bar{k}m^* \\ \left[i \left(i + \int_0^i j + \bar{k}m^* dj + \int_i^{1-\bar{k}m^*} j + \bar{k}m^* dj + \int_{1-\bar{k}m^*}^1 dj \right) - \frac{1}{2}i^2 - \bar{k}m^*(1-i) \right] \\ - \left[i \left(i + \int_0^i j + \bar{k}m^* dj + \int_i^{1-\bar{k}m^*} j + \bar{k}m^* dj + \int_{1-\bar{k}m^*}^1 dj \right) - \frac{1}{2}i^2 - \bar{k}m^*(1-i) \right] = \frac{i^3}{2} & \text{if } i < 1 - \bar{k}m^* \end{cases}$$

The first line refers to a compliant individual. Note that her distrust is increasing in her type. The second line concerns a non-compliant individual. She can only blame those behaving less virtuously than herself. The lower her type the lower her social distrust. It is easy to verify that the first line is always bigger than the second one. Thus, as predicted by Proposition 1.ii, social distrust is higher for more concerned people.

So far we have used a linear one-to-one function between i and $\varepsilon'_i(\cdot)$. By using a steeper (flatter) function and holding average concern constant, one can represent more (less) disagreement in society and verify that total social and institutional distrust are higher (lower). By shifting that function up (down) one can increase (decrease) average concern in society and show that social and institutional distrust increase (decrease). These results would be consistent with Proposition 2. We do not report them here for brevity.

B Proofs

Proof of Lemma 1. For all Group 1's members, the equilibrium behavior satisfies equation (16). The lower bound in this group is the individual who would choose to behave as ρ^* even if she were not fined. Thus her type, $x(\rho^*, m)$, is such that

$$\varepsilon'_{x(\cdot)}(\rho^* + B_{-x(\cdot)}^*(\rho^*, m)) - c'(\rho^*) = 0 \quad (24)$$

If ρ^* and/or m increase, the behavior of affected individuals in equilibrium increases. Thus, by strategic substitutability, all unaffected individuals lower their behavior. Thus (24) is solved by a higher $x(\cdot)$. Moreover, with a higher ρ^* , we would need a higher type $x(\cdot)$ to solve (24) even if all unaffected individuals did not lower their behaviors. 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. For all individuals in Group 2.1, (17) is a strict inequality. They are in a corner solution. The lowest type in this subgroup is $z(\rho^*, m)$ such that

$$\varepsilon'_{z(\cdot)}(\rho^* + B_{-z(\cdot)}^*(\rho^*, m)) - c'(\rho^*) + \frac{\bar{k} \cdot m}{\rho^*} = 0 \quad (25)$$

Type $z(\rho^*, m)$ negatively depends on m . When monitoring is intensified, more individuals will choose to fully comply. Formally, a higher m increases the last term in the LHS of (17), which is positive. So we need a lower $z(\cdot)$ to solve (25).

For all individuals in Group 2.2, (17) holds with equality, so $b_i^*(\rho^*, m) < \rho^*$. By implicit differentiating (17) it follows that $\frac{\partial b_i^*}{\partial m} > 0$. So this group gets smaller as monitoring is intensified, and the behavior of those who “remain” in this group increases. ■

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

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

Take a low-concern individual $l \in L$ and a high-concern individual $h_1 \in H$ with the same level of institutional distrust. By (20), $D_l^I = [\varepsilon_l(\rho_l) - \varepsilon_l(\rho^*)] + [c(\rho^*) - c(\rho_l)] = [\varepsilon_{h_1}(\rho_{h_1}) - \varepsilon_{h_1}(\rho^*)] + [c(\rho^*) - c(\rho_{h_1})] = D_{h_1}^I > 0$. We can re-write as

$$[\varepsilon_{h_1}(\rho_{h_1}) - \varepsilon_{h_1}(\rho^*)] + [\varepsilon_l(\rho^*) - \varepsilon_l(\rho_l)] = [c(\rho^*) - c(\rho_l)] + [c(\rho_{h_1}) + c(\rho^*)].$$

Large convexity of $c(\cdot)$ implies that the last bracketed term in RHS of the equation is quite large. Since it needs to be smaller than the first brackets in the LHS, the latter is large too. Because $\varepsilon_{h_1}(\cdot)$ is quite a concave function we need h_1 to be quite a high type. Thus, if cost function is sufficiently convex and the externality function is sufficiently concave, $|l - a| < |h_1 - a|$. In other words, h_1 is higher than the symmetric type, h . We know that all high-concern types below h_1 distrust the government less than h_1 . Among them, the symmetric type h such that $|l - a| = |h - a|$ distrusts the government less than h_1 and thus less than l . This proves the statement.

- ii. a) Take an unaffected individual i in Group 1, as defined by Lemma 1: $b_i^*(\rho^*, m^*) > \rho^*$.

By (21) and (2),

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

The first squared brackets represents the loss of benefits due to misbehavior of uncompliant people below $z(\rho^*, m^*)$. 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 $\varepsilon_i(\cdot)$, a higher type has to increase his behavior by a larger amount given the misbehavior of uncompliant people.

b) Now consider an affected individual i in Group 2.1, as defined by Lemma 1: $b_i^*(\rho^*, m^*) = \rho^*$. In this case,

$$D_i^S = \left[\varepsilon_i \left(\rho^* + \int_0^{x(\cdot)} \rho^* dj + \int_{x(\cdot)}^1 b_j^*(\rho^*, m^*) dj \right) \right] \quad (26)$$

$$- \left[\varepsilon_i \left(\rho^* + \int_0^{z(\cdot)} b_j^*(\rho^*, m^*) + \int_{z(\cdot)}^{x(\cdot)} \rho^* dj + \int_{x(\cdot)}^1 b_j^*(\rho^*, m^*) dj \right) \right]. \quad (27)$$

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

c) Finally consider a non-compliant individual $i \in [0, z(\cdot))$ in Group 2.2, with $b_i^*(\rho^*, m^*) < \rho^*$ (see Lemma 1). She can only complain that some lower types $j \in [0, i)$ are less non-compliant than she is. This cause the following loss of benefits:

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

In this case, total counterfactual behavior of lower types $j \in [0, i)$ is $\int_0^i b_j^*(\rho^*, m^*) dj$ (i.e., in the counterfactual, all lower types should behave as i does), while their total actual behavior is $\int_0^i b_i^*(\rho^*, m^*) dj$. 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 to that difference. Both factors lead to

bigger social distrust. In addition, the last squared brackets, as in case a), is the cost increase due to the fact that in the counterfactual $b_i^*(\rho^*, m^*)|_{b_j^* \geq b_j^*}$ is lower than i 's actual behavior $b_i^*(\rho^*, m^*)$, 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 disagreement with the dispersion of $\varepsilon'_i(\cdot)$'s, holding the average constant for any values of the argument. With more disagreement low-concern individuals are less concerned and high-concerned are more concerned. The government's policy ρ^* is unchanged. However, both low-concern and high-concern people will distrust the government more because some individual's entitled policy will decrease and increase, respectively. With more disagreement total institutional distrust in society will increase.

As for social distrust, 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 social 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 level of $b_i + B_{-i}$. We can then rewrite eq. (8) at an interior solution as, $\varepsilon'_i(\rho_i^* + B_{-i}^*(\rho_i^*))(1 + \gamma_i) \cdot (1 + x(\rho_i^*)) - c'(\rho_i^*) = 0$. Implicitly differentiating it yields,

$$\frac{\partial \rho_i^*}{\partial \gamma_i} = - \frac{\varepsilon'_i(\cdot) \cdot (1 + x(\cdot))}{\varepsilon''_i(\cdot)(1 + \gamma_i) \cdot (1 + x(\cdot))^2 + \varepsilon'_i(\cdot)(1 + \gamma_i) \cdot x'(\cdot) - c''(\cdot)} > 0,$$

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

Second-order implicit differentiation yields

$$\frac{\partial^2 \rho_i^*}{\partial \gamma_i^2} = \frac{[\varepsilon'_i(\cdot) \cdot (1 + x(\cdot))] \cdot [\varepsilon''_i(\cdot) \cdot (1 + x(\cdot))^2 + \varepsilon'_i(\cdot) \cdot x'(\cdot)]}{\left(\varepsilon''_i(\cdot)(1 + \gamma_i) \cdot (1 + x(\cdot))^2 + \varepsilon'_i(\cdot)(1 + \gamma_i) \cdot x'(\cdot) - c''(\cdot) \right)^2}.$$

The denominator is positive. The first squared bracketed term at the numerator is positive. The second one is positive if $\varepsilon'_i(\rho_i^* + B_{-i}^*(\rho_i^*)) > 0$ 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. Hence, distance between individuals' entitled policy and the government's

policy increases for all individuals. The loss of not implementing the entitled policy is higher for all individuals. Thus total social trust 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 (20) 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 of all individuals whose entitled policy is lower than ρ^* . The increase in ρ^* results in a higher distance and stronger distrust. A second group consists of sufficiently high-concerned individuals, so that the increase in ρ^* results in a lower distance and lower distrust. For some individuals in the third group $\rho_i - \rho^*$ might turn 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. If the increase in ρ^* is sufficiently large, the number of those distrusting the government more is sufficiently large so that total institutional distrust increases.
- ii. By Lemma 1, recall that Equation (17) taken as equality pins down the lowest complying individual, $z(\rho^*)$. Implicitly differentiating Equation (17) yields

$$\frac{\partial z}{\partial \rho^*} = - \frac{\varepsilon_i''(\rho^* + B_{-i}^*(\rho^*, m)) \cdot (1 + \frac{\partial B_{-i}^*(\cdot)}{\partial \rho^*}) - c''(b_i^*(\rho^*, m)) - \frac{\bar{k} \cdot m}{(\rho^*)^2}}{\frac{\partial \varepsilon_i'(\rho^* + B_{-i}^*(\rho^*, m))}{\partial i}} > 0 \quad (28)$$

which implies that the lowest type $z(\rho^*)$ positively depends on ρ^* . In other words, higher ρ^* leads less individuals to fully comply. Moreover, those choosing not to comply reduce their behavior because they are less likely to be caught. This can be shown by implicitly differentiating Equation (17) using the envelope theorem and confirming that implicit derivative of behavior with respect to the government rule is negative, as follows

$$\left. \frac{\partial b_i^*}{\partial \rho^*} \right|_{b_i^* < \rho^*} = - \frac{\varepsilon_i''(b_i^*(\rho^*, m) + B_{-i}^*(\rho^*, m)) \cdot \frac{\partial B_{-i}^*(\cdot)}{\partial \rho^*} - \frac{\bar{k} \cdot m}{(\rho^*)^2}}{\varepsilon_i''(b_i^*(\rho^*, m) + B_{-i}^*(\rho^*, m)) \cdot (1 + \frac{\partial B_{-i}^*(\cdot)}{\partial b_i^*}) - c''(b_i^*(\rho^*, m))} < 0. \quad (29)$$

∂

Intuitively, as ρ^* increases, non-compliant individuals decrease their behavior because the marginal cost of being fined is lower. Those complying will then distrust others more. Any non-compliant individual is more distrustful of those individuals complying less than her because they reduce their behavior.

Both points prove the proposition. ■

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

- i. i) We do not provide a formal proof, but we sketch it. By eq. (8) 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^* distrust the government more because the distance between their entitled policies and ρ^* is bigger. All high-concerned individuals whose entitled policy is higher than ρ^* distrust the government less because the distance between their entitled policies and ρ^* shrinks. The function connecting distrust to types shifts upwards for low-concerned types and downwards 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 that is symmetric with respect to the type preferring ρ^* . Take any high-concerned type h that is symmetric with respect to type l , so that $\rho_h - \rho^* = \rho^* - \rho_l$, 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 $\rho'_h - \rho^* = \rho^* - \rho_{l1} < \rho^* - \rho_l$. 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, $z(\cdot)$, increases). High-types with $b_i^* \geq \rho^*$ will distrust others more. All non-compliant individuals with $b_i^* > \rho^*$ will distrust individuals j with $b_i^* > b_j^*$ because their behavior is lower.

Both points prove the proposition. ■

C 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 2020.⁴⁰

⁴⁰<https://www.respondi.com/EN/>

We collected data from a random sample of adults (above 17 and under 70 years of age) exceeding 2000 individuals per country (see Appendix D.3 for details).⁴¹ 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.⁴² The English questionnaire (link in Appendix E) 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 local surveys in Appendix E).

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.

C.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 given in front of the European Parliament promoting European integration."*⁴³

We then ask the respondents whether they would be willing to read a five-minute long transcription of the speech 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 and the context in which the speech was given was not disclosed to the respondents at the time of choosing. We informed the respondents that their agreement or lack thereof will not affect their payment. In case of agreement,

⁴¹We 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.

⁴²EU-SILC: <https://ec.europa.eu/eurostat/web/main/home>

⁴³For 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>

the respondents were told that they would read and review the text only at the end of the survey.⁴⁴

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 a combination of agreement to read and low-rating assigned to still be interpretable as supportive of the European integration. This would not have been the case with neutral wording, as a respondent antagonising the integration process could have agreed to read the text with the mere intent of assigning a low score. 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 those choosing not to read the text would have been dropped out of the analysis and that the incentivisation would have been lost (it is costly to spend five more 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 behaviours 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.

C.2 The Text Agreement Question: Behavioural Analyses

The analyses here presented follow the analytical framework outlined in Section 3 and confirm the validity of our behavioural 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: $\beta = 19.58$,

⁴⁴For the experimental challenges posed by this question and how they are addressed, see Appendix C.3.

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).⁴⁵ Notice however that though precisely estimated, the coefficient is small: an additional 30 seconds increases the score by 0.02 points.

C.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

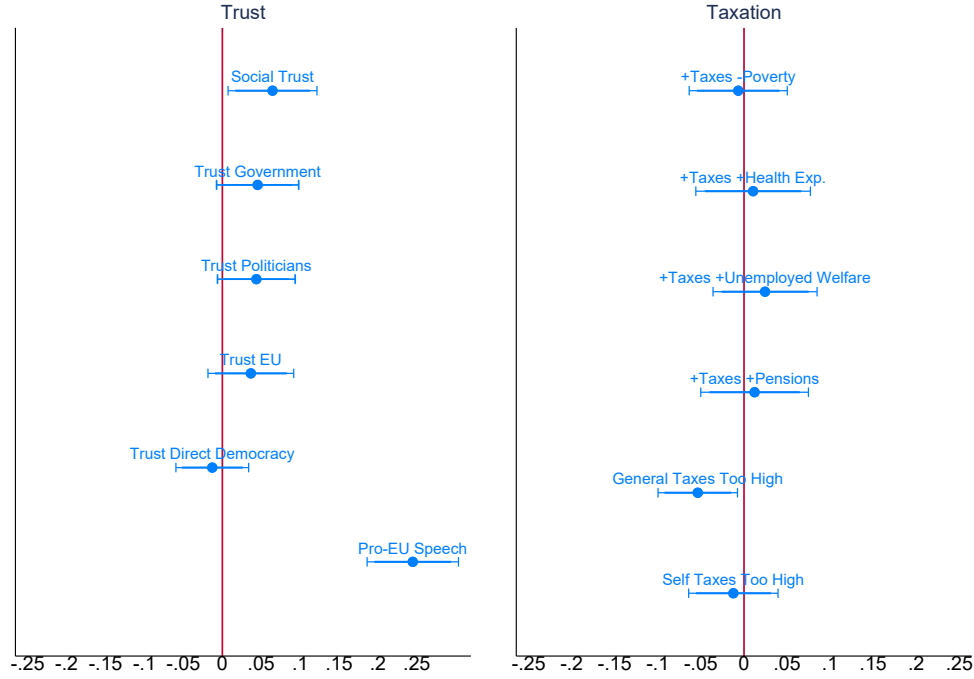
⁴⁵Controls include gender, age groups, employment status, education, immigrant status, family status and number of family members, equivalised household income (coded into five quantiles), a dummy to define the position of the Macron Speech question (see Section C.3 for more details) and regional fixed effects.

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 exploiting 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 C3 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 C1, C2 and C3. 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 C3: 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 C1: Priming Covid and Demand Effects - Health Concerns

	(1) Trust Others	(2) Trust Dir. Democracy	(3) Trust Nat. Gov.	(4) Trust Politicians	(5) Pro EU Speech	(6) Trust EU
Covidfirst	-0.118*** (0.040)	-0.051 (0.031)	-0.007 (0.032)	-0.036 (0.030)	-0.071* (0.038)	-0.135*** (0.033)
Health Concerns	-0.024 (0.022)	0.071*** (0.025)	0.067*** (0.022)	0.068*** (0.022)	0.031** (0.015)	0.109*** (0.023)
Covidfirst*Health Concerns	-0.041* (0.024)	-0.069*** (0.026)	0.058** (0.025)	0.060*** (0.023)	0.038** (0.019)	0.007 (0.029)
Beginning = 1	0.068* (0.041)	-0.008 (0.037)	0.027 (0.031)	0.041 (0.033)	0.392*** (0.037)	0.005 (0.036)
Covidfirst*Beginning	-0.002 (0.050)	-0.007 (0.044)	0.028 (0.043)	0.010 (0.042)	-0.147*** (0.052)	0.038 (0.045)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.079	0.061	0.156	0.214	0.088	0.058
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. 'Health Concerns': self-perceived Covid 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 C2: Priming Covid and Demand Effects - Compliance

	(1) Trust Others	(2) Trust Dir. Democracy	(3) Trust Nat. Gov.	(4) Trust Politicians	(5) Pro EU Speech	(6) Trust EU
Covidfirst	-0.119*** (0.040)	-0.048 (0.031)	-0.004 (0.031)	-0.033 (0.030)	-0.070* (0.038)	-0.130*** (0.033)
Compliance	0.015 (0.021)	0.027 (0.023)	0.098*** (0.022)	0.084*** (0.023)	-0.003 (0.020)	0.120*** (0.024)
Covidfirst*Compliance	-0.064*** (0.024)	-0.047* (0.024)	0.067*** (0.022)	0.050** (0.020)	0.043** (0.022)	0.021 (0.027)
Beginning = 1	0.068* (0.041)	-0.006 (0.037)	0.030 (0.031)	0.044 (0.033)	0.392*** (0.037)	0.010 (0.036)
Covidfirst*Beginning	-0.002 (0.051)	-0.011 (0.044)	0.030 (0.043)	0.010 (0.042)	-0.147*** (0.052)	0.037 (0.045)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.078	0.060	0.162	0.214	0.086	0.061
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 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 C3: Priming Covid and Demand Effects - Pandemic Behaviors

	(1) Trust Others	(2) Trust Dir. Democracy	(3) Trust Nat. Gov.	(4) Trust Politicians	(5) Pro EU Speech	(6) Trust EU
Covidfirst	-0.065 (0.051)	-0.078 (0.055)	0.024 (0.052)	-0.012 (0.047)	-0.147** (0.072)	-0.063 (0.054)
Civic Expectations	0.033 (0.029)	0.020 (0.040)	0.192*** (0.031)	0.140*** (0.033)	-0.036 (0.041)	0.151*** (0.035)
Covidfirst*Civic Expectations	-0.112*** (0.039)	-0.031 (0.047)	0.050 (0.040)	0.029 (0.040)	0.052 (0.048)	0.022 (0.044)
Beginning = 1	0.063 (0.064)	-0.102 (0.070)	-0.012 (0.064)	-0.005 (0.065)	0.269*** (0.077)	-0.010 (0.069)
Covidfirst*Beginning	0.000 (0.085)	0.090 (0.079)	0.026 (0.084)	0.033 (0.077)	-0.004 (0.095)	-0.011 (0.074)
Observations	2,637	2,637	2,637	2,637	2,637	2,637
R-squared	0.118	0.081	0.198	0.241	0.093	0.094
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. 'Civic Expectations': factor variable increasing in expectations of civic behaviour 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

D Tables

Table D4: 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
Power to People	-0.043	1.04	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	0.169	0.375	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 D6: Priming Covid and Trust by Treatment Group

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Power to People	(6) Pro EU Speech
T = 1	-0.101*** (0.029)	0.030 (0.030)	-0.017 (0.029)	-0.120*** (0.035)	-0.058 (0.039)	-0.107*** (0.029)
T = 2	-0.181*** (0.029)	-0.041 (0.029)	-0.082*** (0.028)	-0.187*** (0.034)	-0.032 (0.030)	-0.180*** (0.030)
T = 3	-0.059** (0.030)	0.038 (0.027)	0.011 (0.026)	-0.053* (0.032)	-0.070** (0.031)	-0.140*** (0.032)
Observations	7,916	7,916	7,916	7,916	7,916	7,916
R-squared	0.048	0.147	0.203	0.078	0.060	0.084
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 D7: Priming Covid 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*** (0.023)	-0.072*** (0.024)	-0.058*** (0.022)	-0.072*** (0.022)	0.033 (0.020)	0.059** (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 D8: Priming Covid 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*** (0.029)	-0.048 (0.033)	-0.054* (0.029)	-0.078*** (0.029)	0.011 (0.023)	0.017 (0.030)
T = 2	-0.110*** (0.033)	-0.110*** (0.030)	-0.073** (0.029)	-0.079*** (0.029)	0.076** (0.030)	0.081** (0.039)
T = 3	-0.093*** (0.030)	-0.040 (0.031)	-0.038 (0.029)	-0.056* (0.031)	0.025 (0.026)	0.076*** (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 D9: Priming Covid and Trust - Germany-Netherlands

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Power to People	(6) Pro EU Speech
Covidfirst	-0.129*** (0.029)	-0.005 (0.028)	-0.029 (0.026)	-0.079* (0.041)	-0.061** (0.028)	-0.182*** (0.032)
Observations	4,035	4,035	4,035	4,035	4,035	4,035
R-squared	0.050	0.062	0.055	0.099	0.071	0.070
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	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*** (0.034)	-0.082*** (0.029)	-0.068** (0.033)	-0.087*** (0.031)	0.003 (0.027)	0.052 (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 D10: Priming Covid and Trust - Italy -Spain

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Power to People	(6) Pro EU Speech
Covidfirst	-0.096*** (0.031)	0.022 (0.033)	-0.030 (0.030)	-0.160*** (0.035)	-0.040 (0.039)	-0.104*** (0.029)
Observations	3,881	3,881	3,881	3,881	3,881	3,881
R-squared	0.052	0.074	0.054	0.051	0.042	0.059
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	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** (0.033)	-0.049 (0.037)	-0.038 (0.030)	-0.050* (0.029)	0.078** (0.031)	0.068 (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 D11: Priming Covid, Trust and Compliance

Table D12: Priming Covid, Trust and Civic Expectations

	(1) Trust EU	(2) Trust Nat. Gov.	(3) Trust Politicians	(4) Trust Others	(5) Trust Dir. Democracy	(6) Pro EU Speech
Covidfirst	-0.068* (0.037)	0.036 (0.032)	0.004 (0.033)	-0.065* (0.039)	-0.034 (0.040)	-0.149*** (0.045)
Civic Expectations	0.151*** (0.035)	0.192*** (0.031)	0.140*** (0.033)	0.033 (0.029)	0.019 (0.040)	-0.036 (0.041)
Covidfirst * Civic Expectations	0.022 (0.044)	0.050 (0.040)	0.029 (0.040)	-0.112*** (0.039)	-0.031 (0.047)	0.052 (0.048)
Observations	2,637	2,637	2,637	2,637	2,637	2,637
R-squared	0.094	0.198	0.241	0.118	0.080	0.093
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. ‘Civic Expectations’: factor variable increasing in expectations of civic behaviour 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 D13: Priming Covid, 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*** (0.023)	-0.069*** (0.024)	-0.056** (0.023)	-0.068*** (0.022)	0.034* (0.020)	0.061** (0.026)
Compliance	0.130*** (0.021)	0.184*** (0.023)	0.144*** (0.023)	0.187*** (0.022)	0.065*** (0.020)	0.088*** (0.024)
Covidfirst*Compliance	0.025 (0.024)	-0.016 (0.025)	0.003 (0.027)	-0.026 (0.025)	-0.032 (0.021)	-0.042* (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 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 D14: Priming Covid, Tax and Civic Expectations

	(1)	(2)	(3)	(4)	(5)	(6)
	Tax Poverty	Tax Health Exp.	Tax Unemp. Welf.	Tax Pensions	General Tax	Own Tax
Covidfirst	-0.134*** (0.039)	-0.069* (0.039)	-0.063 (0.039)	-0.046 (0.037)	0.063 (0.044)	0.067 (0.042)
Civic Expectations	0.204*** (0.037)	0.233*** (0.045)	0.159*** (0.038)	0.267*** (0.044)	0.034 (0.044)	0.045 (0.045)
Covidfirst*Civic Expectations	-0.032 (0.046)	-0.019 (0.050)	0.016 (0.048)	-0.076 (0.051)	0.030 (0.052)	0.041 (0.052)
Observations	2,741	2,741	2,741	2,741	2,741	2,741
R-squared	0.069	0.110	0.103	0.105	0.152	0.140
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. ‘Civic Expectations’: factor variable increasing in expectations of civic behaviour 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, equalised 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 D15: Priming Covid, Trust and Health Concerns - Different Countries

	(1)	(2)	(3)	(4)	(5)	(6)
Italy-Spain	Trust Others	Trust Dir. Democracy	Trust Nat. Gov.	Trust Politicians	Pro EU Speech	Trust EU
Covidfirst	-0.156*** (0.036)	-0.026 (0.040)	0.008 (0.034)	-0.047 (0.030)	-0.114*** (0.030)	-0.095*** (0.032)
Health Concerns	-0.020 (0.028)	0.120*** (0.042)	0.020 (0.030)	0.019 (0.023)	0.034 (0.021)	0.116*** (0.033)
Covidfirst* Health Concerns	-0.016 (0.035)	-0.093** (0.041)	0.057 (0.040)	0.074** (0.031)	0.035 (0.028)	-0.033 (0.045)
Observations	3,881	3,881	3,881	3,881	3,881	3,881
R-squared	0.052	0.047	0.078	0.061	0.062	0.059
Controls	YES	YES	YES	YES	YES	YES
Region FE	YES	YES	YES	YES	YES	YES
	(1)	(2)	(3)	(4)	(5)	(6)
Germany-Netherlands	Trust Others	Trust Dir. Democracy	Trust Nat. Gov.	Trust Politicians	Pro EU Speech	Trust EU
Covidfirst	-0.088** (0.042)	-0.071** (0.029)	0.008 (0.029)	-0.018 (0.027)	-0.177*** (0.032)	-0.120*** (0.028)
Health Concerns	-0.035 (0.034)	0.024 (0.032)	0.117*** (0.035)	0.117*** (0.039)	0.037* (0.022)	0.111*** (0.032)
Covidfirst*Health Concerns	-0.044 (0.034)	-0.050 (0.034)	0.058* (0.034)	0.050 (0.035)	0.023 (0.028)	0.040 (0.036)
Observations	4,035	4,035	4,035	4,035	4,035	4,035
R-squared	0.103	0.071	0.087	0.077	0.072	0.069
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. ‘Health Concerns’: self-perceived Covid 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, equalised 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

D.1 Figures

Figure D4: Distribution of the Health Concerns Variable

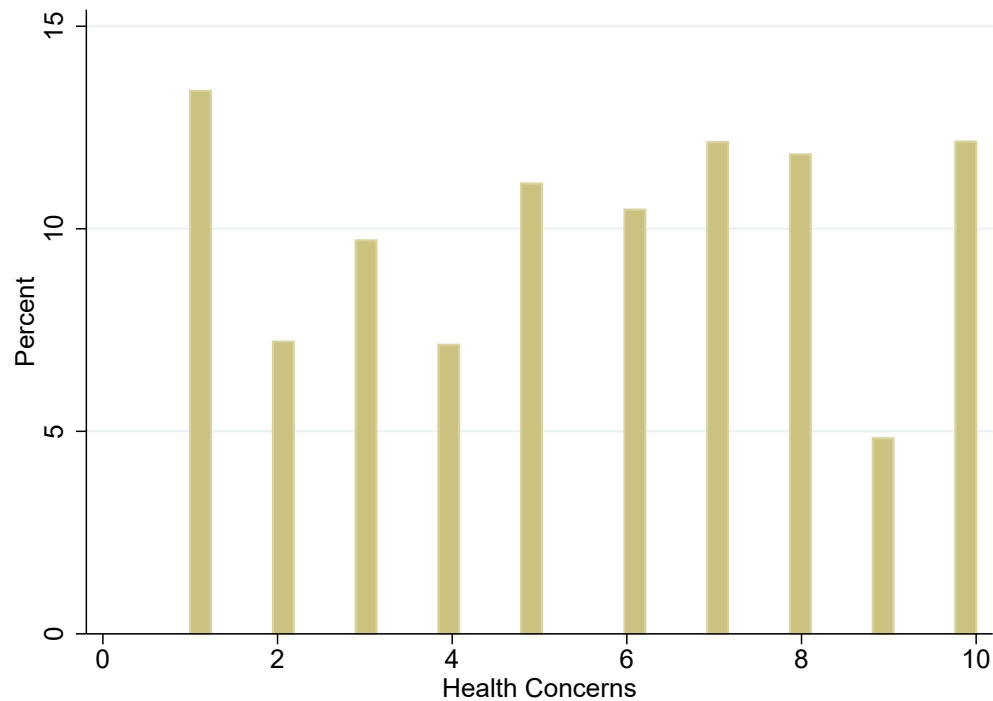


Figure D5: The figures displays the distribution of the variable Health Concerns.

Figure D6: Placebo: Interaction Term

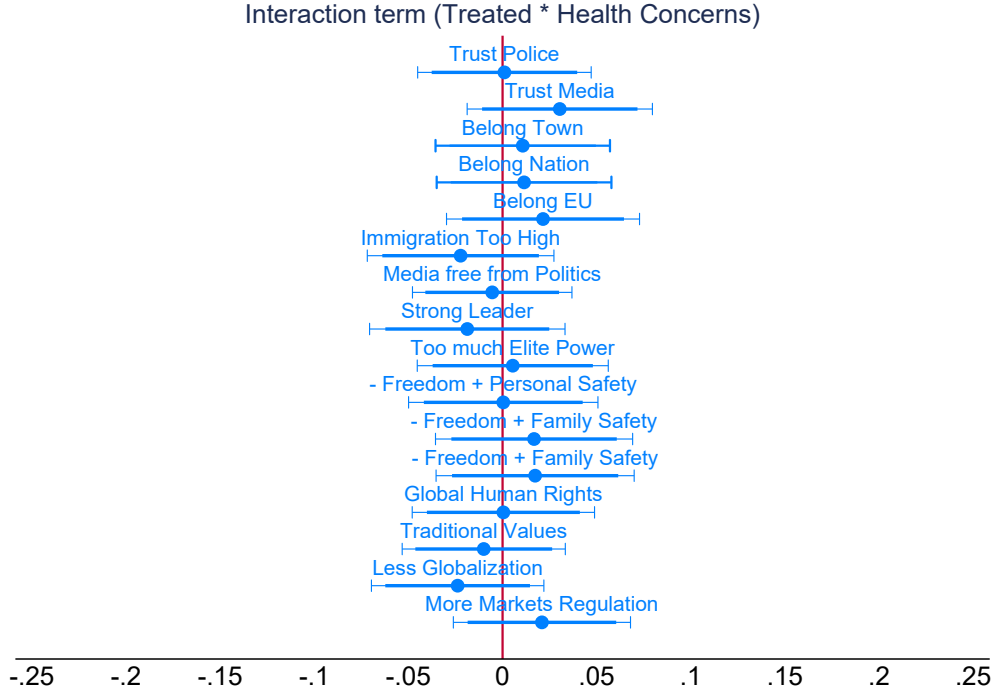


Figure D7: The figures display the interaction term COVIDFIRST * Health Concerns on an alternative set of outcomes included in the original survey. From the top: trust in police, trust in media, attachment to your town/country/European Union, immigration is too low (0) or too high (10) in your country, 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. Check Appendix E for the complete list of questions in the survey.

D.2 Multiple Hypothesis & Placebo Tests

We correct the p-values of our coefficients for the multiplicity of tested hypotheses (List et al., 2019; Barsbai et al., 2020). We report the uncorrected p-values in Table D16, as well as the family-wise corrections, i.e. trust, attitudes towards the European Union, voting behaviours and taxation. All our significant estimates survive, apart from trust in politicians. Finally, the last column of Table D16 reports the p-values corrected for the simultaneous estimation of all the equations for which we have presented the results. Except for trust in politicians, demand for devolution of power to the people, and the perception of one's own tax burden, all results remain within conventional significance levels.⁴⁶

⁴⁶Notice that for computational reasons we are not able to apply clustering of the observations in our correction.

Table D16: Uncorrected p-values and p-values corrected for multiplicity of hypotheses

	Uncorrected	Family-wise	All equations
Trust			
Trust Government	0.8239	0.8239	0.976
Trust Politicians	0.095*	0.135	0.277
Social Trust	< 0.001***	< 0.002***	< 0.001***
Trust EU	< 0.001***	< 0.001***	0.002***
Pro-EU Speech	< 0.001***	< 0.001***	0.003***
Trust Direct Democracy	0.024**	0.071*	0.09*
Taxation and welfare state contributions			
+ Taxes - Poverty	< 0.001***	< 0.001***	< 0.001***
+ Taxes + Health expenditure	0.003***	0.012**	0.033**
+ Taxes + Unemployment Welfare	0.019***	0.024**	0.097*
+ Taxes + Pensions	0.003***	0.014**	0.033**
General Tax Too High	0.188	0.188	0.620
Own Tax Too High	0.011***	0.031**	0.101

Significance stars denote conventional significance levels.

D.3 Sample size and balancing test

Tables [D17](#) and [D18](#) report respectively the sample size for each country and the balance of our sample across conditions.

Table D17: 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 D18: Balancing Test

Variables	Δ : Covidfirst-Baseline	Δ : Health-Baseline	Δ : Economic-Baseline	Δ : Conflict-Baseline
Unemployed	-0.012** (0.041)	-0.012 (0.104)	-0.014* (0.064)	-0.009 (0.238)
Education	0.076** (0.029)	0.083* (0.065)	0.059 (0.187)	0.084* (0.058)
Native	-0.000 (0.998)	-0.002 (0.804)	0.003 (0.676)	-0.001 (0.864)
Female	-0.018 (0.127)	-0.031** (0.039)	-0.014 (0.355)	-0.008 (0.574)
Age	-0.002 (0.834)	0.011 (0.335)	-0.009 (0.428)	-0.008 (0.498)
Household size	0.032 (0.160)	0.017 (0.578)	0.051* (0.080)	0.029 (0.320)
HH income	-0.021 (0.271)	-0.015 (0.525)	-0.005 (0.853)	-0.042* (0.082)
Single	0.013 (0.233)	0.024* (0.092)	0.006 (0.692)	0.010 (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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

E 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_850cx81c4806tzT

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_0ln902bfxiBsH1r