Moral dilemmas and trust in leaders during a global health crisis

Jim A.C. Everett[†1], Clara Colombatto[†2], Edmond Awad[3], Paulo Boggio[4], Björn Bos[5], William J. Brady[2], Megha Chawla[2], Vladimir Chituc[2], Dongil Chung[6], Moritz A. Drupp[5], Srishti Goel[2], Brit Grosskopf[3], Frederik Hjorth[7], Alissa Ji[2], Caleb Kealoha[2], Yangfei Lin[3], Yina Ma[8,9], Michel A. Maréchal[10], Federico Mancinelli[11], Christoph Mathys[11,12,13], Asmus L. Olsen[7], Graeme Pearce[3], Annayah M.B. Prosser[14], Niv Reggev[15], Nicholas Sabin[16], Julien Senn[10], Walter Sinnott-Armstrong[17], Hallgeir Sjåstad[18], Madelijn Strick[19], Sunhae Sul[20], Lars Tummers[21], Monique Turner[22], Hongbo Yu[23], Yoonseo Zoh[2], Molly J. Crockett*[2]

[†]equal contribution

*Correspondence to: molly.crockett@yale.edu

[1] School of Psychology, University of Kent, Canterbury, UK
[2] Department of Psychology, Yale University, New Haven, CT, USA
[3] Department of Economics, University of Exeter, Exeter, UK
[4] Social and Cognitive Neuroscience Laboratory, Mackenzie Presbyterian University, São Paulo, Brasil
[5] Department of Economics, University of Hamburg, Hamburg, Germany
[6] Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea
[7] Department of Political Science, University of Copenhagen, Copenhagen, Denmark
[8] State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China
[9] Chinese Institute for Brain Research, Beijing, China
[10] Department of Economics, University of Zurich, Zurich, Switzerland
[11] Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
[12] Interacting Minds Centre, Aarhus University, Aarhus, Denmark
[13] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland
[14] Department of Psychology, University of Bath, Bath, UK
[15] Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[16] Department of Management, Faculty of Management and Economics, Universidad de Santiago de Chile, Santiago, Chile.
[17] Department of Philosophy, Duke University, Durham, NC, USA
[18] Department of Strategy and Management, Norwegian School of Economics, Bergen, Norway
[19] Department of Psychology, Utrecht University, Utrecht, The Netherlands
[20] Department of Psychology, Pusan National University, Busan, South Korea
[21] School of Governance, Utrecht University, Utrecht, The Netherlands
[22] Department of Communication, Michigan State University, East Lansing, MI, USA
[23] Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA, USA

Abstract

Trust in leaders is central to citizen compliance with public policies. One potential determinant of trust is how leaders resolve conflicts between utilitarian and non-utilitarian ethical principles in moral dilemmas. Past research suggests utilitarian responses to dilemmas can both erode and enhance trust in leaders: sacrificing some people to save many others (i.e., instrumental harm) reduces trust, while impartially maximizing the welfare of everyone equally (i.e., impartial beneficence) may increase trust. Here, we investigate moral dilemmas and trust in leaders during the COVID-19 pandemic. Across 22 countries on 6 continents, participants will complete behavioral and self-reported measures of trust in leaders who endorse either utilitarian or non-utilitarian principles in COVID-19 dilemmas. We predict that endorsement of instrumental harm will decrease trust, while endorsement of impartial beneficence will increase trust. These results can advance our understanding of trust in leaders and inform effective public communication during times of crisis.

Introduction

During times of crisis, such as wars, natural disasters, or pandemics, citizens look to leaders for guidance. Successful crisis management often depends on mobilizing individual citizens to change their behaviors and make personal sacrifices for the public good [1]. Crucial to this endeavour is trust: citizens are more likely to follow official guidance when they trust their leaders [2]. Here, we investigate public trust in leaders in the context of the COVID-19 pandemic, which continues to threaten millions of lives around the globe at the time of writing [3,4].

Because the novel coronavirus is highly transmissive, a critical factor in limiting pandemic spread is compliance with public health recommendations such as social distancing, physical hygiene and mask wearing [5,6]. Trust in leaders is a strong predictor of citizen compliance with a variety of public health policies [7–12]. During pandemics, trust in experts issuing public health guidelines is a key predictor of compliance with those guidelines. For example, during the avian influenza pandemic of 2009 (H1N1), self-reported trust in medical organizations predicted self-reported compliance with protective health measures and vaccination rates [13,14]. During the COVID-19 pandemic, data from several countries shows that public trust in scientists, doctors, and the government is positively associated with self-reported compliance with public health recommendations [15,16,17,18] . These data suggest that trust in leaders is likely to be a key predictor of long-term success in containing the COVID-19 pandemic around the globe. However, the factors that *determine* trust in leaders during global crises remain understudied.

One possible determinant of trust in leaders during a crisis is how they resolve moral dilemmas that pit distinct ethical principles against one another. The COVID-19 pandemic has

raised particularly stark dilemmas of this kind: for instance, whether to prioritize young and otherwise healthy people over the elderly and people with chronic illnesses when allocating scarce medical treatments [19,20]. This dilemma and similar others highlight a tension between two major approaches to ethics. Consequentialist theories – of which utilitarianism is the most well-known exemplar [21] – posit that only consequences should matter when making moral decisions. Because younger, healthier people are more likely to recover and have longer lives ahead of them, utilitarians would argue that they should be prioritized for care because this is likely to produce the best overall consequences [22–24]. In contrast, non-utilitarian theories of morality, such as deontological theories [25–29], argue that morality should consider more than just consequences, including rights, duties, and obligations (see Supplementary Note 1 for further details). Non-utilitarians, on deontological grounds, could argue that everyone who is eligible (e.g. by being a citizen and/or contributing through taxes or private health insurance) has an equal right to receive medical care, and therefore it is wrong to prioritize some over others [30]. While it is unlikely that ordinary citizens *explicitly* think about moral issues in terms of specific ethical theories [21,31], past work shows these philosophical concepts explain substantial variance in the moral judgments of ordinary citizens [32,33], including in the context of the COVID-19 pandemic[34].

There is robust evidence that people who endorse utilitarian principles in sacrificial dilemmas – deeming it morally acceptable to sacrifice some lives to save many others – are seen as less moral and trustworthy, chosen less frequently as social partners, and trusted less in economic exchanges than people who take a non-utilitarian position and reject sacrificing some to save many [35–40]. This suggests that leaders who take a utilitarian approach to COVID-19 dilemmas will be trusted less than leaders who take a non-utilitarian approach. Anecdotally, some recent case studies of public communications are consistent with this hypothesis. In the US, for example, public discussions around whether to reopen schools and the economy versus remain in lockdown highlighted tensions between utilitarian approaches and other ethical principles, with some leaders stressing an imperative to remain in lockdown to prevent deaths from COVID-19 (consistent with deontological principles), and others arguing that lockdown also has costs and these need to be weighed against the costs of pandemic-related deaths (consistent with utilitarian principles; see Supplementary Note 2). Those who appealed to utilitarian arguments – such as President Donald Trump, who argued "we cannot let the cure be worse than the problem itself" [41] and Texas Lieutenant Governor Dan Patrick, who suggested that elderly Americans might be "willing to take a chance" on their survival for the sake of their grandchildrens' economic prospects [42] – were met with widespread public outrage [43]. Likewise, when leaders in Italy suggested prioritizing young and healthy COVID-19 patients over the elderly when ventilators became scarce, they were intensely criticized by the public [44]. Mandatory contact tracing policies, which have been proposed on utilitarian grounds, have also faced strong public criticisms about infringement on individual rights to privacy [45–47].

While past research and recent case studies suggest that utilitarian approaches to pandemic dilemmas are likely to erode trust in leaders, other evidence suggests this conclusion may be premature. First, some work shows that utilitarians are perceived as more competent than

non-utilitarians [38] , and to the extent that trust in leaders is related to perceptions of their competence [2], it is possible that utilitarian approaches to pandemic dilemmas will increase rather than decrease trust in leaders. Second, utilitarianism has at least two distinct dimensions: it permits harming innocent individuals to maximize aggregate utility (instrumental harm), and it treats the interests of all individuals as equally important (impartial beneficence)[21,33]. Indeed, preliminary evidence suggests these two dimensions characterize the way ordinary people think about moral dilemmas in the context of the COVID-19 pandemic [34]. These two dimensions of utilitarianism are not only psychologically distinct in the general public[33], but they also have distinct impacts on perception of leaders. Specifically, utilitarians are seen as *worse* political leaders when they endorse instrumental harm, but in some cases, they are seen as *better* political leaders when they endorse impartial beneficence[37].

Another dilemma that pits utilitarian principles against other non-utilitarian principles – this time in the domain of impartial beneficence – is whether leaders should prioritize their own citizens over people in other countries when allocating scarce resources. The utilitarian sole focus on consequences mandates a strict form of impartiality: the mere fact that someone is one's friend (or their mother, or their fellow citizen) does not imply that they have any obligations to such a person that they do not have to any and all persons [48]. Faced with a decision about whether to help a friend (or family member, or fellow citizen) or instead provide an equal-or-slightly-larger benefit to a stranger, this strict utilitarian impartiality means that one cannot morally justify favouring the person closer to them. In contrast, many non-utilitarian approaches explicitly incorporate these notions of special obligations, recognising the relationships between people as morally significant. Here, President Trump went against utilitarian principles when he ordered a major company developing personal protective equipment (PPE) to stop distributing it to other countries who needed it[49], or when he ordered the U.S. government to buy up all the global stocks of the COVID-19 treatment Remdesivir [50]. His actions generated outrage across the world and stood in contrast to statements from many other Western leaders at the time. The Prime Minister of the UK, Boris Johnson, for example, endorsed impartial beneficence when he argued for the imperative to "ensure that the world's poorest countries have the support they need to slow the spread of the virus"[51]. In a similar vein, the Dutch government donated 50 Million Euro to CEPI, an organization that aims to distribute vaccines equally across the world [52].

In sum, public trust in leaders is likely to be a crucial determinant of successful pandemic response, and may depend in part on how leaders approach the many moral dilemmas that arise during a pandemic. Utilitarian responses to such dilemmas may erode or enhance trust relative to non-utilitarian approaches, depending on whether they concern instrumental harm or impartial beneficence. Past research on trust and utilitarianism is insufficient to understand how utilitarian resolutions to moral dilemmas influence trust during the COVID-19 pandemic – and future crises – for several reasons. First, it has relied on highly artificial moral dilemmas, such as the "trolley problem" [53,54], that most people have not encountered in their daily lives. Thus, the findings of past studies may not generalize to the context of a global health crisis, where everyone around

the world is directly impacted by the moral dilemmas that arise during a pandemic. Second, because the vast majority of previous work on trust in utilitarians has focused on instrumental harm, we know little about how impartial beneficence impacts trust. Third, most previous work on this topic has focused on trust in ordinary people. However, there is evidence that utilitarianism differentially impacts perceptions of ordinary people and leaders [37,38,40], which means we cannot generalize from past research on trust in utilitarians to a leadership context. Because leaders have power to resolve moral dilemmas through policymaking, and therefore can have far more impact on the outcomes of public health crises than ordinary people can, it is especially important to understand how leaders' approaches to moral dilemmas impact trust. Finally, past work on inferring trust from moral decisions has been conducted in just a handful of Western populations – the US, UK, and Germany – and so may not generalize to other countries that are also affected by the COVID-19 pandemic. We need, therefore, to assess cross-cultural stability by testing this hypothesis in different countries around the world. Indeed, given observations of cultural variation in the willingness to endorse sacrificial harm[32], it is not a foregone conclusion that utilitarian decisions will impact trust in leaders universally. For further details of how the present work advances our understanding of moral dilemmas and trust in leaders, please see Supplementary Notes 3-5.

The goal of the current research is to test the hypothesis that endorsement of instrumental harm will decrease trust in leaders, while endorsement of impartial beneficence will increase trust in leaders in the context of the COVID-19 pandemic. Testing this hypothesis across a diverse set of 22 countries spanning six continents (Figure 1A) may inform how leaders around the globe can communicate with their constituencies in ways that will preserve trust during global crises. Given the public health consequences of mistrust in leaders [7–9], if our hypothesis is confirmed, leaders may wish to carefully consider weighing in publicly on moral dilemmas that are unresolvable with policy, because their opinions might erode citizens' trust in other pronouncements that may be more pressing, such as advice to comply with public health guidelines.

To test our hypothesis empirically, we draw on case studies of public communications to identify five moral dilemmas that have been actively debated during the COVID-19 pandemic (Figure 1B). Three of these dilemmas involve instrumental harm: the *Ventilators* dilemma concerns whether younger individuals should be prioritized to receive intensive medical care over older individuals when medical resources such as ventilators are scarce [44,55], the *Lockdown* dilemma concerns whether to consider reopening schools and the economy or remain in lockdown [55,56], and the *Tracing* dilemma concerns whether it should be mandatory for residents to carry devices that continuously trace the wearer's movements, allowing the government to immediately identify people who have potentially been exposed to the coronavirus [45–47]. The other two dilemmas involve impartial beneficence: the *PPE* dilemma concerns whether PPE manufactured within a particular country should be reserved for that country's citizens under conditions of scarcity, or sent where it is most needed [55,57–59], and the *Medicine* dilemma concerns whether a novel COVID-19 treatment developed within a particular country should be

delivered with priority to that country's citizens, or shared impartially around the world [57,60,61]. Participants in our studies will read about leaders who endorse either utilitarian or non-utilitarian solutions to the dilemmas, and subsequently complete behavioral and self-reported measures of trust in the respective leaders. For example, they may read about a leader who endorsed prioritizing the young over the elderly for scarce ventilators, and then be asked how much they trust that leader. While there are many similar dilemmas potentially relevant to the COVID-19 crisis, we chose to focus on the five described above because they (1) have been publicly debated at time of writing; and (2) apply to all countries in our planned sample. For further details of why we chose these specific dilemmas and how they can test our theoretical predictions, please see Supplementary Notes 2 and 6-9.

We will measure trust in two complementary ways. First, we will ask participants to self-report their general trust in the leaders, both in terms of an overall character judgment ("How trustworthy do you think this person is?") as well as how likely they would be to trust this person on other issues not related to the dilemma ("How likely would you be to trust this person's advice on other issues?"). Second, we will use a novel, incentivized *voting task* designed to measure public trust in leaders (Figure 1C). Following past work, we define leaders as people who are responsible for making decisions on behalf of a group[62]. In the voting task, participants are invited to cast a vote to appoint a leader who is responsible for making a charitable donation on behalf of a group. Crucially, the leader has the opportunity to "embezzle" some of the donation money for themselves. Participants will be asked to vote for either a person who endorsed a utilitarian or a non-utilitarian position on a COVID-19 dilemma; the person who receives the most votes will have control over the group's donation. By measuring preferences for a leader who will be responsible for a group's donations to help those in need, the voting task captures trust in leaders in a specific context that is highly relevant to our central research question: during a health crisis, effective leadership requires responsible stewardship of public resources in order to help those in need. For further details of why we designed our trust measures in this way, please see Supplementary Notes 10-12.

Our analyses will thus test two complementary hypotheses. First, we predict that self-reported trust will be lower for leaders who endorse utilitarian over non-utilitarian approaches to dilemmas involving instrumental harm, while the reverse pattern will be observed for dilemmas involving impartial beneficence (Hypothesis 1). Second, we predict that participants will be less likely to vote for leaders who endorse utilitarian over non-utilitarian views on dilemmas involving instrumental harm, while the reverse pattern will be observed for dilemmas involving impartial beneficence (Hypothesis 2). Pilot studies conducted in the US and UK in July 2020 provided initial support for these hypotheses (see Pilot Data in Supplementary Information for details). All analyses will control for participants' demographics and own policy preferences in each dilemma.

Finally, we note that the framing of both the self-reported and behavioral measures of trust are deliberately unrelated to the pandemic dilemmas we use to highlight the moral commitments of the leader. This crucial design choice will allow us to measure the impact of

utilitarian versus non-utilitarian endorsements of pandemic dilemmas on *subsequent trust* in leaders. In this way, the current design illuminates an important real-life question: if a leader weighs in publicly on a moral dilemma during a crisis, how likely are they to be trusted later on other matters of public concern?

Methods

Ethics Information

Our research complies with all relevant ethical regulations. The study was approved by the Yale Human Research Protection Program Institutional Review Board (Protocol IDs: 2000027892 and 2000022385), the Ben-Gurion University of the Negev Human Subjects Research Committee (Request Number 20TrustCovR), the Centre for Experimental Social Sciences Ethics Committee (OE_0055), and the NHH Norwegian School of Economics Institutional Review Board (NHH-IRB 10/20). Informed consent will be obtained from all participants.

Design

Overview. An overview of the experiment is depicted in Figure 2. After selecting their language, providing their consent, and passing two attention checks, participants will be told that they will "read about three different debates that are happening right now around the world", that they will be given "some of the justifications that politicians and experts are giving for different policies", and that they will be "ask[ed] some questions about [their] opinions". They will then complete two tasks measuring their trust in leaders expressing either utilitarian or non-utilitarian opinions (one using a behavioral measure and one using self-report measures, presented in a randomized order); these tasks will be followed by questions about their impressions about the ongoing pandemic crisis, as well as individual difference and demographic measures, as detailed below. Data collection will be performed blind to the conditions of the participants.

Both behavioral and self-report measures of trust will involve five debates on the current pandemic crisis, three of which involve instrumental harm (IH), and two impartial beneficence (IB) (summarized in Figure 1B and Table 2; for full text, see Supplementary Methods). Each of these five dilemmas are based on real debates that have been occurring during the COVID-19 pandemic, and we developed the philosophical components of each argument in consultation with moral philosophers.

1) **Lockdown** (Instrumental Harm): whether the country should maintain severe restrictions on social gatherings until a vaccine is developed to prevent COVID-related deaths, or consider relaxing restrictions to maximize overall well-being
2) **Ventilators** (Instrumental Harm): whether doctors should give everyone equal access to COVID treatment, or prioritize younger and healthier people

3) **Tracing** (Instrumental Harm): whether the government should make it mandatory for residents to wear contact tracing devices to prevent pandemic spread, or make tracing devices optional to respect residents' right to privacy
4) **Medicine** (Impartial Beneficence): whether medicine developed in the home country should be reserved for treating the home country's citizens, or sent wherever it can do the most good, even if that means sending it to other countries
5) **PPE** (Impartial Beneficence): whether PPE manufactured in the home country should be reserved for protecting the home country's citizens, or sent wherever it can do the most good, even if that means sending it to other countries

See Supplementary Notes 2 and 6-9 for further details of why we chose these specific dilemmas and how they can test our theoretical predictions.

Translations. Where the survey is being conducted in a non-English speaking country, study materials will be translated following a standard forward- and back-translation procedure[68]. First, for forward translation, a native speaker will translate materials from English to the target language. Second, for back translation, a second naive translator (who has not seen the original English materials) will translate the materials back into English. Results will be compared and if there are any substantial discrepancies, a second forward- and back-translation will be conducted with translators working in tandem to resolve issues. Finally, the finished translated and back-translated materials will be checked by researchers coordinating the experiment for that country.

Experimental design. Participants will be randomly and blindly assigned to one of 4 conditions in the beginning of the experiment. These conditions correspond to a 2x2 between-subjects design: 2 (moral dimension in the voting task: Instrumental Harm/Impartial Beneficence) x 2 (argument in the self-reported trust task: Utilitarian/Non-Utilitarian). In addition, we will randomize the order of tasks (voting or self-reported trust first), the order of arguments in the voting task (Utilitarian or Non-Utilitarian first), the order of dilemmas in the self-reported trust (Lockdown, Ventilators, or Tracing first if Instrumental Harm, and PPE or Medicine first if Impartial Beneficence), and the dilemmas displayed (two in the self-reported trust task and one in the voting task randomly chosen among Lockdown, Ventilators, and Tracing if Instrumental Harm, and PPE and Medicine if Impartial Beneficence). This design allows us to minimize demand characteristics with between-subjects manipulations of key experimental factors, while at the same time maximizing efficiency of data collection.

Attention checks. We have two attention checks prior to the beginning of the experiment. Any participants who fail one or both of these are then screened out immediately. First, participants will be told:

> "In studies like ours, there are sometimes a few people who do not carefully read the questions they are asked and just "quickly click through the survey." These random

answers are problematic because they compromise the results of the studies. It is very important that you pay attention and read each question. In order to show that you read our questions carefully (and regardless of your own opinion), please answer "Tik Tok" in the question on the next page"

Then, on the next page, participants will be given a decoy question: "When an important event is happening or is about to happen, many people try to get informed about the development of the situation. In such situations, where do you get your information from?". They will be asked to select among the following possible answers, displayed in a randomized order: Tik Tok, TV, Twitter, Radio, Reddit, Facebook, Youtube, Newspapers, Other. Participants who fail to follow our instructions and select any answer other than the instructed one ("Tik Tok") are then screened out of the survey.

Second, participants will be asked to read a short paragraph about the history and geography of roses. On the following page, they will be asked to indicate which of six topics was not discussed in the paragraph. Participants who answer incorrectly are then screened out of the survey.

*Dilemma introduction***.** Both the voting and self-report trust tasks will begin with an introduction to a specific dilemma. In the voting task, participants view a single dilemma, and in the self-report task, participants view two dilemmas in randomized order (see Figure 2 for details). No participant will see the same dilemma in both voting and self-report tasks.

The dilemma introduction will consist of a short description of the dilemma (e.g. in the "PPE" dilemma: "Imagine that [...] there will soon be another global shortage of personal protective equipment [... and] political leaders are debating how personal protective equipment should be distributed around the globe."), followed by a description of two potential policies (e.g. in the "PPE" dilemma, American participants will read: "Some are arguing that PPE made in American factories should be given to whoever needs it most, even if that means sending it to other countries. Others are arguing that PPE made in American factories should be kept in the U.S., because the government should focus on protecting its own citizens.").

After reading about the dilemma, participants will be asked to provide their own opinion about the best course of action ("Which policy do you think should be adopted?"), to be answered on a 1-7 scale, with the endpoints (1 and 7) representing strong preferences for one of the policies (e.g. in the "PPE" dilemma they will be labeled "Strongly support U.S.-made PPE being reserved for protecting American citizens" and "Strongly support U.S.-made PPE being given to whoever needs most", respectively), and the midpoint (4) representing indifference ("Indifferent"). See Supplementary Note 13 for further details. As an exploratory measure that will not be analyzed for the purposes of the current report, participants will also indicate how morally wrong it would be for politicians to endorse the utilitarian approach in each dilemma.

For full text of dilemmas and introduction questions, see Supplementary Methods.

Voting task. Our behavioral measure of trust in the current studies is based on a novel task with two types of participants: *voters* and *donors.* Voters are asked to cast a vote for a leader who is responsible for making a charitable donation to UNICEF on behalf of a group of donors, and has the opportunity to "embezzle" some of the donation money for themselves (Figure 1C).

We will collect data from donors first. A few days before we run our main experiment, a convenience sample of American participants (N = 100) will be recruited from Prolific and be provided with a $2 bonus endowment. They will be given the opportunity to donate up to their full bonus to UNICEF. After making their donation decision, they will read about the five COVID-19 dilemmas, in randomized order, and indicate which policy they think should be adopted. Finally, they will be instructed that they may be selected to be responsible for the entire group's donations to UNICEF. If they are selected, they will have the opportunity to keep up to the full amount of total group donations for themselves. They will be asked to indicate how much of the group's donations they would keep for themselves if they are selected to be responsible.

Our main experiment focuses on the behavior of voter participants. In the voting task, participants will be randomly assigned to read about one dilemma, randomly selected amongst the five dilemmas summarised in Table 2. After completing the dilemma introduction, participants will be asked to "make a choice that has real financial consequences". They will be told that "[a] few days ago, a group of 100 people were recruited via an international online marketplace and invited to make donations to the charitable organization UNICEF. In total, they donated an amount equivalent to [donation amount in local currency]". We will instruct participants that we would like them to "vote for a leader to be responsible for the entire group's donations". Crucially, they will also be told that "[t]he leader has two options: They can transfer the group's [donation amount] donation to UNICEF in full, *or* they can take some of this money for themselves (up to the full amount) and transfer whatever amount is left to UNICEF". The exact donation amount will be determined by the actual donation choices of the donor participants.

Following these details, participants will be asked to cast a vote for the leadership position between two people who had also read about the same dilemma they had just read about. Participants will be instructed that one person agreed with the utilitarian argument, and the other person agreed with the non-utilitarian argument. This information will be displayed to participants on the same page, in a randomized order. Participants will then be asked to vote for the person they wished to be responsible for the group's donations. We will instruct participants that we will later identify the winner of the election, and implement their choice by distributing payments to the leader and UNICEF accordingly.

After completing the voting task, voter participants will be asked the following comprehension question: "On the last page, you were asked to choose a leader that will be entrusted with the group's donation. Please select the option that best describes what the leader will be able to do with the donation". They will be asked to select between three options, displayed in randomized order:

1. The leader chooses how much of the group's donation to keep for themselves and how much to transfer to UNICEF.
2. The leader chooses how much of the group's donation to transfer to The Red Cross and how much to transfer to Medecins Sans Frontiers.
3. The leader chooses how much of the group's donation to keep for themselves and how much to return to the people who donated the money.

We will exclude voter participants who fail to select the correct answer (1), as per our exclusion criteria (see Exclusions).

After collecting the votes from the voter participants, we will randomly select ten donor participants to be considered for the leadership position: one who endorsed the utilitarian position for each of the five dilemmas and one who endorsed the non-utilitarian position for each of the five dilemmas. After tallying the votes from voter participants, we will implement the choices of each of the elected leaders.

For full text of instructions and questions for both the donor and the voting task, see Supplementary Methods.

Self-reported trust. Participants will read about two dilemmas on the dimension of utilitarianism that they will *not* encounter in the voting task. Participants assigned to an Instrumental Harm dilemma (Lockdown, Ventilators, or Tracing) for the voting task will read both Impartial Beneficence dilemmas (PPE and Medicine) for the self-reported trust task, while participants assigned to an Impartial Beneficence dilemma (PPE or Medicine) for the voting task will read a randomly assigned two out of three Instrumental Harm dilemmas (Lockdown, Ventilators, and Tracing) for the self-reported trust task. The structure of the introduction to the dilemmas will be identical to that in the voting task: they will read a short description of the issue, followed by a description of two potential policies. On separate screens, they will be asked which policy they themselves support.

After providing their own opinions, participants will be asked to imagine that the mayor of a major city in their region is arguing for one of the two policies, providing either a utilitarian or non-utilitarian argument. Each participant will be randomly assigned to read about leaders making either utilitarian or non-utilitarian arguments in all dilemmas presented in the self-report trust task. After reading about the leader's opinion and argument, they will then be asked to report their general trust in the leader ("How trustworthy do you think this person is?"), to be answered on a 1-7 scale, with labels "Not at all trustworthy", "Somewhat trustworthy", and "Extremely trustworthy", at points 1, 4, 7, respectively. On a separate page they will then be asked to report their trust in the leader's advice on other issues ("How likely would you be to trust this person's advice on other issues?"), to be answered on a 1-7 scale, with labels "Not at all likely", "Somewhat likely", and "Extremely likely", at points 1, 4, 7, respectively.

After completing the self-reported trust task, participants will be asked the following comprehension question: "On the last page, you read about a mayor in a city in your region, and

were asked about them. Please select the option that best describes the questions you were asked". Their options, displayed in a randomized order, will be: (1) How much I agreed with the mayor; (2) How much I trusted the mayor; and (3) How much I admired the mayor. This will allow us to exclude participants who fail to select the correct answer (2), as per our exclusion criteria (see Exclusions).

For full text of instructions and questions for the self-report trust task, see Supplementary Methods.

COVID concern. To assess their attitudes toward and experience with the pandemic, participants will be asked three questions. Two will measure how concerned participants currently feel about the pandemic, on both health-related and economic grounds ("How concerned are you about the health-related consequences of the COVID-19 pandemic?", and "How concerned are you about the financial and economic consequences of the COVID-19 pandemic?", both to be answered on a 1-7 scale, with labels "Not at all", and "Very much", at points 1 and 7, respectively). The third question will measure their personal involvement ("Have you or anyone you know personally suffered significant health consequences as a result of COVID-19?", to be answered by selecting one of three options: "Yes", "No", "Unsure").

Oxford Utilitarianism Scale. All participants will then complete the Oxford Utilitarianism Scale[33]. The scale consists of nine items in two subscales: instrumental harm (OUS-IH) and impartial beneficence (OUS-IB). The OUS-IB subscale consists of five items that measure endorsement of impartial maximization of the greater good, even at great personal cost (e.g., "It is morally wrong to keep money that one doesn't really need if one can donate it to causes that provide effective help to those who will benefit a great deal"). The OUS-IH subscale consists of four items relating to willingness to cause harm so as to bring about the greater good (e.g.,"It is morally right to harm an innocent person if harming them is a necessary means to helping several other innocent people"). Participants will view all questions in a randomized order, and answer on a 1-7 scale, with labels "Strongly disagree", "Disagree", "Somewhat disagree", "Neither agree nor disagree", "Somewhat agree", "Agree", and "Strongly agree".

Demographics. All participants will be asked to report their gender, age, years spent in education, subjective SES, education (on the same scale, but with minor changes in the scale labels across countries), political ideology (using an item from the World Values Survey), and religiosity. These questions will be the same across countries and represent the demographics used as covariates in the main analysis. Additionally, participants will be asked to indicate their region of residence (e.g. for US, "Which US State do you currently live in?"), and ethnicity/race, with the specific wording and response options depending on the local context. In addition, participants will be asked to confirm their country of residence, which will allow us to exclude participants who report living in a different country than that of intended recruitment, as per our exclusion criteria (see Exclusions).

Debrief questions. Finally, participants will be asked a series of debriefing questions. Two of these will be assessing their participation in other COVID-related studies ("Approximately how many COVID-related surveys have you participated in before this one?", to be answered by selecting one of the following options: "0", "1-5", "6-10", "11-20", "21-50", "more than 50", "I don't remember"; and "If you have participated in any other COVID-related studies, how similar were they to this one?", to be answered by selecting one of the following options: "Extremely similar", "Very similar", "Moderately similar", "Slightly similar", "Not at all similar", "Not applicable").

An additional question will assess participants' attitudes towards the charity involved in the voting task ("How reliable do you think UNICEF is as an organization in using donations for helping people?", to be answered on a 1-5 scale, with labels "Not reliable at all", "Somewhat reliable", and "Very reliable", at points 1, 3, and 5, respectively).

Analysis Plan

Pre-processing

Exclusions. Data will be excluded either at the participant level as outlined in the Analysis Plan section, based on criteria 1 (duplicate response), 2 (different residence), and 3 (partial completion), or on an analysis-by-analysis basis as outlined in criteria 4 (missing variables) and 5 (failed comprehension checks).

Outliers. All participants' responses will be analyzed.

Computation of composite measures. Composite measures of self-reported trust will be created by averaging responses to the two trust questions (trustworthiness of the leader, and trust in the leader's advice on other issues), separately for each participant and dilemma. In addition, we will create composite OUS scores for each participant by averaging their responses on the scale items, separately for the Instrumental Harm (4 items) and Impartial Beneficence sub-scales (5 items).

Analysis plan for hypothesis testing

We will examine behavioral measures and self-reported measures of trust in two separate models. For testing our hypotheses across all countries, we will set a significance threshold of alpha = 0.0025 (Bonferroni corrected for two tests). All analyses will be conducted in R using packages lmer [69], lmerTest [70], and emmeans [71]. In the event of convergence or singularity issues, we will supplement the theoretically appropriate models described below with simplified models by reducing complexity of the random effects structure [72].

Hypothesis 1: Self-reported trust. To examine participants' self-reported trust in the leaders, we will examine the composite measure of their trust in each leader (i.e. the average of the two trust questions, computed separately for each participant and dilemma). We hypothesize that participants will report higher trust in non-utilitarian leaders compared to utilitarian leaders in the context of dilemmas involving instrumental harm, while the opposite pattern will be observed for impartial beneficence. To test this hypothesis, we will conduct a linear mixed-effects model of the effect of argument type (Utilitarian vs. Non-Utilitarian), dimension type (Instrumental Harm vs. Impartial Beneficence), and their interaction, on the composite score of trust, adding demographic variables (namely gender, age, education, subjective SES, political ideology, and religiosity) and policy support as fixed effects, and dilemmas and countries as random intercepts, with participants nested within countries. In addition, we will run a model that includes countries as random slopes of the two main effects and the interactive effect. Should the model converge, and should the results differ from the simpler model proposed above, we will compare model fits using the Akaike information criterion (AIC), and retain the model that better fits the data - while still reporting the other in supplementary materials. We will follow up on significant effects with post-hoc comparisons using Bonferroni corrections. For the purposes of the analysis, we will use effect coding such that for argument type, the Non-Utilitarian condition will be coded as -0.5 and the Utilitarian condition as 0.5, and for the dimension type, Instrumental Harm will be coded as -0.5, and Impartial Beneficence as 0.5. The demographic covariates will be grand mean-centered; the gender variable will be dummy coded with "female" as baseline. P-values will be computed using Satterthwaite's approximation for degrees of freedom as implemented in lmerTest. For analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

Hypothesis 2: Voting measure. To examine participants' trust in the leaders as demonstrated by their behavior, we will examine their choices in the voting task, where they will be asked to select which of two leaders (one making a utilitarian argument, and the other a non-utilitarian one) to entrust with a group charity donation. We hypothesize that participants will be more likely to select the non-utilitarian leader over the utilitarian leader when reading about their arguments for dilemmas involving instrumental harm, while the opposite pattern will be observed for impartial beneficence. To test this hypothesis, we will conduct a generalized linear mixed-effects model with the logit link of the effect of dimension type (Instrumental Harm vs. Impartial Beneficence) on the leader choice (Utilitarian vs. Non-Utilitarian), adding demographic variables (namely gender, age, education, subjective SES, political ideology, and religiosity) and policy support as fixed effects, and dilemmas and countries as random intercepts, with participants nested within countries. In addition, we will also run a model that includes countries as random slopes of the effect of dimension type. Should the model converge, and should the results differ from the simpler model proposed above, we will compare model fits using the Akaike information criterion (AIC), and retain the model that better fits the data - while still

reporting the other in supplementary materials. Based on recent reports that linear models might be preferable to logistic models in treatment designs[73,74], we will run the same analysis using a linear model (instead of logit link) with the identical fixed and random effects and again adjudicate between the models using the AIC. We will follow up on significant effects with post-hoc comparisons using Bonferroni corrections. For the purposes of this analysis, we will use effect coding such that for the binary response variable of argument type, the Non-Utilitarian trust response will be coded as 0 and the Utilitarian trust response as 1, and for the dimension type, Instrumental Harm will be coded as -0.5, and Impartial Beneficence as 0.5. The demographic covariates will be grand mean-centered; the gender variable will be dummy coded with "female" as baseline. P-values will be computed using Satterthwaite's approximation for degrees of freedom as implemented in lmerTest. For analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

Robustness checks. Because there is evidence that public perceptions of lockdowns are changing relative to July 2020 when we ran our pilots [75,76], which may affect responses to the Lockdown dilemma, we will examine the robustness of our findings using two variations of the models described above, one that includes the Lockdown dilemma and another that omits the dilemma.

As some of the countries in our sample already implement mandatory and/or invasive contact tracing schemes at the time of writing (China, India, Israel, Singapore, and South Korea), which may affect responses to the Tracing dilemma, we will examine the robustness of our findings in these countries using two variations of the models described above, one that includes the Tracing dilemma and another that omits the dilemma. Furthermore, in this subset of countries we will examine an order effect to test whether completing the Tracing dilemma in the first task affects behavior on the subsequent task.

Null Hypothesis Testing. In the event of non-significant results from the approaches outlined above, we will employ the Two One-Sided Tests (TOST) procedure[63] to differentiate between insensitive vs. null results. In particular, we will specify lower and upper equivalence bounds based on standardized effect sizes set by our SESOI (see Power Analysis section, or Table 1). For each of our two tasks, should the larger of the two *p* values from the two t-tests be smaller than alpha = .05, we will conclude statistical equivalence. For example, the minimum guaranteed sample size (N = 12600; see Sample Size for details) will give us over 95% power to detect an effect size of $d = 0.05$ in the self-reported trust task, yielding standardized $\Delta_L = -0.05$ and $\Delta_U = 0.05$, and $OR = 1.30$ in the voting task, yielding standardized $\Delta_L = -0.15$ and $\Delta_U = 0.15$.

Sampling Plan

Participants. The study will be completed online by participants in the following countries: Australia, Brasil, Canada, Chile, China, Denmark, France, Germany, India, Israel, Italy, the Kingdom of Saudi Arabia, Mexico, the Netherlands, Norway, Singapore, South Africa,

South Korea, Spain, United Arab Emirates, United Kingdom, and United States of America (Figure 1A). We sample on every inhabited continent and include countries that have been more and less severely affected by COVID-19 on a variety of metrics (see Supplementary Figure 1). Country selection was determined primarily on a convenience basis. In April 2020 the senior author put out a call for collaborators via social media and email. Potential collaborators were asked whether they had the capacity to recruit up to 1,000 participants representative for age and gender within their home country. After the initial set of collaborators was established, we added additional countries to diversify our sample with respect to geographic location and pandemic severity.

Participants will be recruited via online survey platforms (see Supplementary Table 1) and compensated financially for their participation in accordance with local standard rates. We will aim to recruit samples that are nationally representative with respect to age and gender where feasible. We anticipate this will be feasible for many, but not all countries in our study (see Supplementary Table 1 for details). Sampling will take place over a 14-day period. All survey materials will be translated to the local language (see Translations for details). Prior to the survey, all participants will read and approve a consent form outlining their risks and benefits, confirm they agree to participate in the experiment, and complete an attention check. Participants who fail to agree to the consent or fail to pass the attention check will not be permitted to complete the survey.

Expected effect sizes. We can inform our expected effect sizes from examining the published literature on utilitarianism and trust. Previous studies of social impressions of utilitarians reveal effect sizes in the range of $d = 0.19$ - 0.78 (mean $d = 0.78$ for the effect of instrumental harm on self-reported moral impressions; mean $d = 0.19$ for the effect of impartial beneficence on self-reported moral impressions; mean $d = 0.55$ for interactive effects of instrumental harm and impartial beneficence on self-reported moral impressions)[35–40]. However, there are several important caveats with using these past studies to inform expected effect sizes for the current study. First, past studies have measured trust in ordinary people, while we study trust in leaders, and there is evidence that instrumental harm and impartial beneficence differentially impact attitudes about leaders versus ordinary people[37]. Second, past studies have investigated artificial moral dilemmas, while we study real moral dilemmas in the context of an ongoing pandemic. Third, past studies have been conducted in a small number of Western countries (US, UK, Germany) while we sample across a much wider range of countries on 6 continents. Finally, for the voting task, it is more challenging to estimate an expected effect size because no previous studies to our knowledge have used such a task.

Because of the caveats described above, we also informed our expectations of effect sizes with data from Pilot 2, which was identical to the proposed studies in design apart from using The Red Cross instead of UNICEF in the voting task and the omission of the Tracing dilemma (see Pilot Data in Supplementary Information for a full description of the pilot experiments). Pilot 2 revealed a conventionally medium effect size for the interaction between argument and

moral dimension in the self-reported trust task (B = 2.88, SE = 0.24, $t(452)$ = 11.80, $p < .001$, CI = [2.41, 3.35], d = 0.55), and a conventionally large effect size for the effect of moral dimension in the voting task (B = 2.41, SE = 0.33, z = 7.30, $p < .001$, CI = [1.77, 3.13], OR = 11.13, d = 1.33).

Sample size. Sample size was determined based on a cost-benefit analysis considering available resources and expected effect sizes that would be theoretically informative[63] (see Expected effect sizes). We aimed to collect the largest sample possible with resources available and verified with power analyses that our planned sample would be able to detect effect sizes that are theoretically informative and at least as large as expected based on prior literature (see Power analysis). At the time of writing, we expect to collect a sample of 21,000 participants in total. Conservatively accounting for exclusion rates up to 40% (see Exclusions), this would lead to a final guaranteed minimum sample of 12,600 participants.

Power analysis. We conducted a series of power analyses to determine the smallest effect sizes that our minimum guaranteed sample of 12,600 participants would be able to detect with 95% power and an alpha level of .005, separately for each main model (see Analysis plan for further details). To account for these two hypothesis tests, for all power analyses we applied Bonferroni corrections for two tests, thus yielding an alpha of 0.0025. Following recent suggestions [64, cf.65], results passing a corrected alpha of $p \leq 0.005$ will be interpreted as '*supportive* evidence' for our hypotheses, while results passing a corrected alpha of $p < 0.05$ will be interpreted as '*suggestive* evidence'. Power analyses were conducted using Monte Carlo simulations[66] via the R package simr[67], with 1000 simulations, using estimates of means and variances from Pilot 2 (see Pilot Data in Supplementary Information for a full description of the pilot experiments; note that for the purposes of the current simulations, the race variable was omitted from data analysis because this variable is not readily comparable across countries). Data and code for power analyses can be found at https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

First, we considered the interactive effect of moral dimension (Instrumental Harm vs. Impartial Beneficence) and argument (Utilitarian vs. Non-Utilitarian) on trust in the self-report task. We estimate that a sample of 12,600 participants will provide over 95% power to detect an effect size of d = .05 (power = 99.3%, CI = [98.56, 99.72]). This effect size is 9% of what we observed in Pilot 2 and is the smallest effect size of interest (SESOI) for the self-report task.

Next, we considered the effect of moral dimension (Instrumental Harm vs. Impartial Beneficence) on leader choice in the voting task. We estimate that a sample of 12,600 participants will provide over 95% power to detect an odds ratio of 1.30 (power = 95.8%, CI = [94.36, 96.96]). This effect size is 9% of what we observed in Pilot 2 and is the SESOI for the voting task.

Given that these smallest effect sizes of interest are detectable at 95% power with our guaranteed sample (total N = 12,600), are theoretically informative, and are lower than our

expected effect sizes (see Expected effect sizes), we conclude that our sample is sufficient to provide over 95% power for testing our hypotheses, and that our study is highly powered to detect useful effects.

We note that at time of writing, online survey platform representatives indicated that while it is normally feasible to recruit samples nationally representative for age and gender in most of our target countries, due to the ongoing pandemic, final sample sizes may be unpredictable and in some countries it will not be possible to achieve fully representative quotas for some demographic categories, including women and older people (see Supplementary Table 1 for details). If this issue arises, we will prioritize statistical power over representativeness. If we are unable to achieve representativeness for age and/or gender in particular countries, we will note this explicitly in the results section.

Exclusions. Participants will be excluded from all further analyses if they meet at least one of the following criteria: (1) they have taken the survey more than once (as indicated by IP address or worker ID); (2) they report in a question about their residence (further described in Design) that they live in a different country than that of intended recruitment; (3) they do not answer more than 50% of the questions. In addition, participants will be selectively excluded from specific analyses if they (4) do not provide a response and are thus missing variables involved in the analysis; or (5) fail the comprehension check (further described in Design) for the task involved in the analysis.

Data availability statement

All data and materials will be made openly available on the Open Science Framework (OSF) website at this link: https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

Code availability statement

All analysis code (completed in R) will be made openly available on the Open Science Framework (OSF) website at this link: https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

References

1. Wilson, S. Pandemic leadership: Lessons from New Zealand's approach to COVID-19. *Leadership* **16**, 279–293 (2020).
2. Levi, M. & Stoker, L. Political Trust and Trustworthiness. *Annu. Rev. Polit. Sci.* **3**, 475–507 (2000).
3. Ferguson, N. *et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. http://spiral.imperial.ac.uk/handle/10044/1/77482 (2020)

doi:10.25561/77482.
4. Fink, S. Worst-Case Estimates for U.S. Coronavirus Deaths - The New York Times. *The New York Times* (2020).
5. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
6. Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
7. Alsan, M. & Wanamaker, M. Tuskegee and the Health of Black Men*. *Q. J. Econ.* **133**, 407–455 (2018).
8. Christensen, D., Dube, O., Haushofer, J., Siddiqi, B. & Voors, M. BUILDING RESILIENT HEALTH SYSTEMS: 110 (2020).
9. Lowes, S. & Montero, E. The Legacy of Colonial Medicine in Central Africa. 80 (2020).
10. Udow-Phillips, M. & Lantz, P. Trust in Public Health Is Essential Amid the COVID-19 Pandemic. *J. Hosp. Med.* **15**, 431–433 (2020).
11. Blair, R. A., Morse, B. S. & Tsai, L. L. Public health and public trust: Survey evidence from the Ebola Virus Disease epidemic in Liberia. *Soc. Sci. Med. 1982* **172**, 89–97 (2017).
12. Rubin, G. J., Amlôt, R., Page, L. & Wessely, S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. *BMJ* **339**, (2009).
13. Gilles, I. *et al.* Trust in medical organizations predicts pandemic (H1N1) 2009 vaccination behavior and perceived efficacy of protection measures in the Swiss public. *Eur. J. Epidemiol.* **26**, 203–210 (2011).
14. Prati, G., Pietrantoni, L. & Zani, B. Compliance with recommendations for pandemic influenza H1N1 2009: the role of trust and personal beliefs. *Health Educ. Res.* **26**, 761–769 (2011).
15. Maher, P. J., MacCarron, P. & Quayle, M. Mapping public health responses with attitude networks: the emergence of opinion□based groups in the UK's early COVID□19 response phase. *Br. J. Soc. Psychol.* **59**, 641–652 (2020).
16. Plohl, N. & Musil, B. Modeling compliance with COVID-19 prevention guidelines: the critical role of trust in science. *Psychol. Health Med.* 1–12 (2020) doi:10.1080/13548506.2020.1772988.
17. Dohle, S., Wingen, T. & Schreiber, M. *Acceptance and Adoption of Protective Measures During the COVID-19 Pandemic: The Role of Trust in Politics and Trust in Science*. https://osf.io/w52nv (2020) doi:10.31219/osf.io/w52nv.
18. Han, Q. *et al. Trust in government and its associations with health behaviour and prosocial behaviour during the COVID-19 pandemic*. https://osf.io/p5gns (2020) doi:10.31234/osf.io/p5gns.
19. Bramble, B. *Pandemic Ethics: 8 Big Questions of COVID-19*. (Bartleby Books, 2020).
20. Emanuel, E. J. *et al.* Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N. Engl. J. Med.* **382**, 2049–2055 (2020).
21. Everett, J. A. C. & Kahane, G. Switching Tracks? Towards a Multidimensional Model of Utilitarian Psychology. *Trends Cogn. Sci.* **24**, 124–134 (2020).
22. Giubilini, A., Savulescu, J. & Wilkinson, D. COVID-19 Vaccine: Vaccinate the Young to Protect the Old? *J. Law Biosci.* (2020).
23. Savulescu, J., Persson, I. & Wilkinson, D. Utilitarianism and the pandemic. *Bioethics* **34**, 620–632 (2020).
24. Savulescu, J. & Cameron, J. Why lockdown of the elderly is not ageist and why levelling down equality is wrong. *J. Med. Ethics* medethics-2020-106336 (2020) doi:10.1136/medethics-2020-106336.
25. Fried, C. *Right and Wrong*. (Harvard University Press, 1978).
26. Kant, I. *Groundwork for the Metaphysics of Morals*. (Yale University Press, 2002).
27. Rawls, J. *A Theory of Justice*. (Belknap Press of Harvard University Press, 1971).
28. Ross, W. D. *The right and the good*. (Oxford University Press, 1930).
29. Scanlon, T. *What we owe to each other*. (Belknap Press, 1998).
30. Liddell, K., Martin, S. & Palmer, S. Allocating Medical Resources in the Time of Covid-19. *N. Engl.*

J. Med. **382**, e79 (2020).
31. Conway, P., Goldstein-Greenwood, J., Polacek, D. & Greene, J. D. Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition* **179**, 241–265 (2018).
32. Awad, E., Dsouza, S., Shariff, A., Rahwan, I. & Bonnefon, J.-F. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proc. Natl. Acad. Sci.* **117**, 2332–2337 (2020).
33. Kahane, Everett, G. *et al.* Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychol. Rev.* **125**, 131–164 (2018).
34. Navajas, J. *et al. Utilitarian Reasoning About Moral Problems of the COVID-19 Crisis*. https://osf.io/ktv6z (2020) doi:10.31219/osf.io/ktv6z.
35. Bostyn, D. H. & Roets, A. Trust, trolleys and social dilemmas: A replication study. *J. Exp. Psychol. Gen.* **146**, e1–e7 (2017).
36. Everett, J. A. C., Pizarro, D. A. & Crockett, M. J. Inference of Trustworthiness From Intuitive Moral Judgments. *J. Exp. Psychol. Gen.* **145**, 772–787 (2016).
37. Everett, E., Jim A. C., Faber, N. S., Savulescu, J. & Crockett, M. J. The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *J. Exp. Soc. Psychol.* **79**, 200–216 (2018).
38. Rom, S. C., Weiss, A. & Conway, P. Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *J. Exp. Soc. Psychol.* **69**, 44–58 (2017).
39. Sacco, D. F., Brown, M., Lustgraaf, C. J. N. & Hugenberg, K. The Adaptive Utility of Deontology: Deontological Moral Decision-Making Fosters Perceptions of Trust and Likeability. *Evol. Psychol. Sci.* **3**, 125–132 (2017).
40. Uhlmann, E. L., Zhu, L. (Lei) & Tannenbaum, D. When it takes a bad person to do the right thing. *Cognition* **126**, 326–334 (2013).
41. Trump, D. J. WE CANNOT LET THE CURE BE WORSE THAN THE PROBLEM ITSELF. AT THE END OF THE 15 DAY PERIOD, WE WILL MAKE A DECISION AS TO WHICH WAY WE WANT TO GO! *Twitter* https://twitter.com/realDonaldTrump/status/1241935285916782593?s=20 (2020).
42. Patrick, D. Tucker Carlson Tonight. (2020).
43. Burke, D. Reopening the country: The dangerous moral arguments behind this movement - CNN. https://edition.cnn.com/2020/04/23/us/reopening-country-coronavirus-utilitarianism/index.html (2020).
44. Rosenbaum, L. Facing Covid-19 in Italy — Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. *N. Engl. J. Med.* **382**, 1873–1875 (2020).
45. Fahey, R. A. & Hino, A. COVID-19, digital privacy, and the social limits on data-focused public health responses. *Int. J. Inf. Manag.* (2020) doi:10.1016/j.ijinfomgt.2020.102181.
46. Asher. TraceTogether: Singapore turns to wearable contact-tracing Covid tech. *BBC News* (2020).
47. From India to Cyprus, understanding the global debate over virus contact tracing apps. *The Week* https://www.theweek.in/news/world/2020/05/07/understanding-the-global-privacy-debate-over-coronavirus-contact-tracing-apps.html (2020).
48. Jeske, D. Special Obligations. in *The Stanford Encyclopedia of Philosophy* (ed. Zalta, E. N.) (Metaphysics Research Lab, Stanford University, 2014).
49. Breuninger, K. & Wilkie, C. Trump bans export of coronavirus protection gear, says he's 'not happy with 3M'. *CNBC* https://www.cnbc.com/2020/04/03/coronavirus-trump-to-ban-export-of-protective-gear-after-slamming-3m.html (2020).
50. U.S. Department of Health and Human Services. Trump Administration Secures New Supplies of Remdesivir for the United States. *HHS.gov* https://www.hhs.gov/about/news/2020/06/29/trump-administration-secures-new-supplies-remdesivir-united-states.html (2020).
51. Boris, J. Prime Minister's statement on coronavirus (COVID-19): 3 June 2020. *GOV.UK* https://www.gov.uk/government/speeches/pm-statement-at-the-coronavirus-press-conference-3-june-2020.

52. Kerris, M. ‘Onze missie: de hele wereld een vaccin’. *NRC Handelsblad* https://www.nrc.nl/nieuws/2020/05/14/onze-missie-de-hele-wereld-een-vaccin-a3999818 (2020).
53. Foot, P. The Problem of Abortion and the Doctrine of the Double Effect. *Oxf. Rev.* **5**, 5–15 (1967).
54. Thomson, J. J. The Trolley Problem. *Yale Law J.* **94**, 1395–1415 (1985).
55. Savulescu, J., Persson, I. & Wilkinson, D. Utilitarianism and the pandemic. *Bioethics* **34**, 620–632 (2020).
56. Kupferschmidt, K. The lockdowns worked—but what comes next? *Science* **368**, 218–219 (2020).
57. Gertz, G. Coordinating the international distribution of medical goods. in *Reopening the world: How to save lives and livelihoods* (eds. Allen, J. R. & West, D. M.) 12–15 (The Brookings Institution, 2020).
58. Mehrotra, P., Malani, P. & Yadav, P. Personal Protective Equipment Shortages During COVID-19—Supply Chain–Related Causes and Mitigation Strategies. *JAMA Health Forum* **1**, e200553–e200553 (2020).
59. Zhou, Y. R. The global effort to tackle the coronavirus face mask shortage. *The Conversation* http://theconversation.com/the-global-effort-to-tackle-the-coronavirus-face-mask-shortage-133656 (2020).
60. Bollyky, T. J., Gostin, L. O. & Hamburg, M. A. The Equitable Distribution of COVID-19 Therapeutics and Vaccines. *JAMA* **323**, 2462–2463 (2020).
61. Liu, Y., Salwi, S. & Drolet, B. C. Multivalue ethical framework for fair global allocation of a COVID-19 vaccine. *J. Med. Ethics* **46**, 499–501 (2020).
62. Edelson, M. G., Polania, R., Ruff, C. C., Fehr, E. & Hare, T. A. Computational and neurobiological foundations of leadership decisions. *Science* **361**, eaat0036 (2018).
63. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence Testing for Psychological Research: A Tutorial. *Adv. Methods Pract. Psychol. Sci.* **1**, 259–269 (2018).
64. Benjamin, D. J. *et al.* Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2018).
65. Lakens, D. *et al.* Justify your alpha. *Nat. Hum. Behav.* **2**, 168–171 (2018).
66. Brysbaert, M. & Stevens, M. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *J. Cogn.* **1**, 9 (2018).
67. Green, P. & MacLeod, C. J. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* **7**, 493–498 (2016).
68. Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M. & Zamir, S. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *Am. Econ. Rev.* **81**, 1068–1095 (1991).
69. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *ArXiv14065823 Stat* (2014).
70. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **82**, (2017).
71. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. Emmeans: Estimated marginal means, aka least-squares means. *R Package Version* **1**, 3 (2018).
72. Barr, D. J. Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* **4**, (2013).
73. Gomila, R. Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *J. Exp. Psychol. Gen.* (2020) doi:10.1037/xge0000920.
74. Angrist, J. D. & Pischke, J. S. *Mostly harmless econometrics: An empiricist’s companion.* (Princeton University Press., 2009).
75. Bosman, J., Mervosh, S. & Santora, M. As the Coronavirus Surges, a New Culprit Emerges: Pandemic Fatigue. *The New York Times* (2020).
76. Santora, M. & Kwai, I. As Virus Surges in Europe, Resistance to New Restrictions Also Grows. *The New York Times* (2020).
77. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).

876

Acknowledgments

Pilot data collection was supported by Prolific Academic. JACE was partly supported through the British Academy, Leverhulme Trust, and the Department for Business, Energy, and Industrial Strategy (SRG19\190050). WJB was supported by a National Science Foundation fellowship (award #1808868). DC was partly supported by the National Research Foundation of Korea (NRF-2018R1D1A1B0704358). VC was supported by a National Science Foundation Graduate Research Fellowship. EA, BG, YL and GP thank the University of Exeter Business School for funding their contribution to this research. BB and MAD are grateful for support from central internationalization funds of the Universität Hamburg. NS gratefully acknowledges funding support provided by the Department of Management, Faculty of Management and Economics, Universidad de Santiago de Chile, and ANID FONDECYT de Iniciación en Investigación 2020 (Folio 11200781). ALO and FH gratefully acknowledge support from the Independent Research Fund Denmark (0213-00052B) and the Faculty of the Social Sciences at the University of Copenhagen. AMBP was supported by a scholarship from the Economic & Social Research Council. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

We thank Joshua Monrad for advice on scenario design; Junior Okoroafor, Douglas Shao and Xinran Wang for assistance; and Johannes Apel, Aline Bidani, Nina Breedveld, Rachel Calcott, Ryan Carlson, Leonardo Alfaro Cui cui, Andrés A. Gálvez, Judy Kim, Nahrie Kim, Felicia Michelsen, Mikkel Meinert Pedersen, Aviv Mokady, Astrid Oline Ervik, Jiaxin Yang, Xiaoyu Zeng, and Michele Zoccali for assistance with survey translations.

Author Contributions

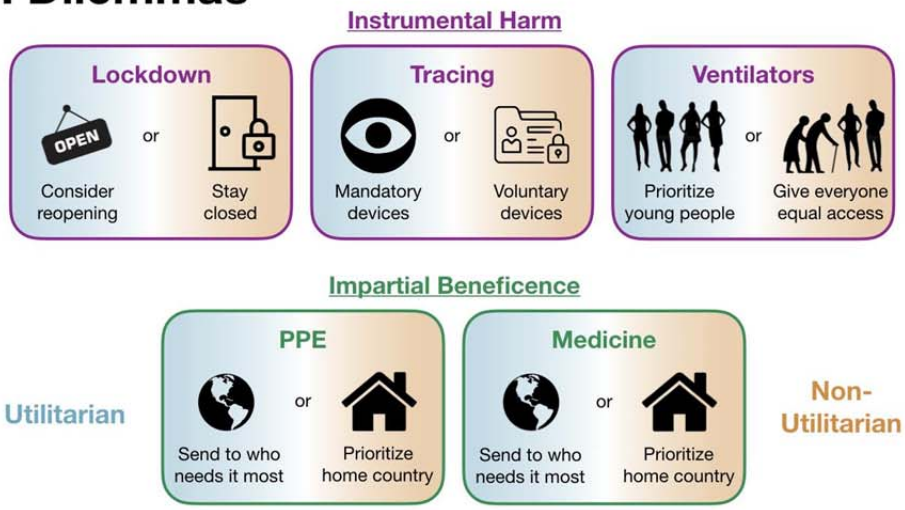
MJC, JACE, CC, VC, and WJB conceived the research. MJC, JACE, CC, EA, PB, BB, WJB, MC, VC, DC, MAD, SG, FH, YM, MAM, CM, ALO, AMBP, NR, NS, JS, WSA, HS, MS, SS, LT, MT, HY, & YZ designed the research. MAM, JS, MJC, JACE, CC, HS, LT, NS, and EA developed the voting task. JACE, VC, MJC, CC, and WSA wrote the moral dilemmas. CC and JE analyzed the pilot data in consultation with MJC, WJB, AJ, CM, and NR. CC conducted the power analysis in consultation with MJC, WJB, CM and NR. CC, JACE, MJC, WJB, CM, and NR developed the analysis plan. JACE, CC, and MJC prepared the manuscript with feedback from all co-authors. MJC, JACE, CC, and CK coordinated the implementation of the project. MJC, JACE, CC, EA, PB, BB, MC, DC, MAD, SG, BG, FH, CK, YL, YM, MAM, FM, CM, ALO, GP, NR, NS, JS, HS, MS, SS, LT, HY, and YZ will contribute to data collection and/or translation. All co-authors will review and approve the final manuscript.

915 ***Competing Interests***
916
917 The authors declare no competing interests
918

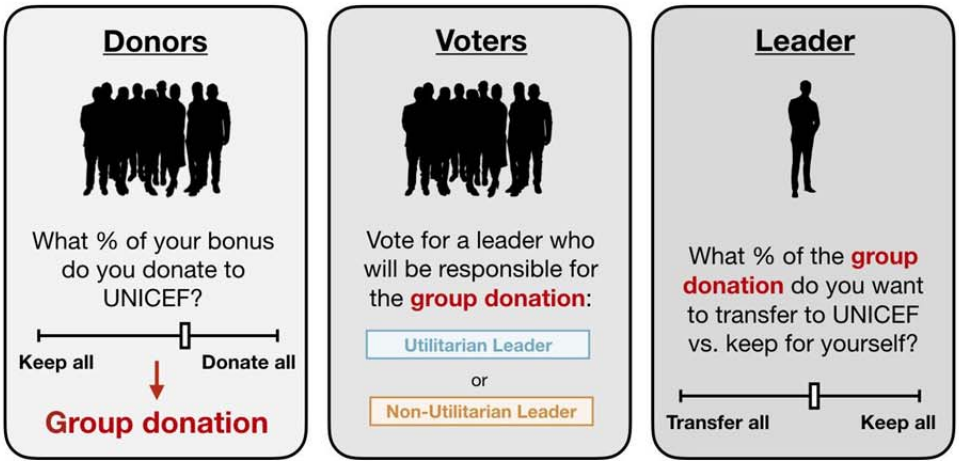
Figures and Figure Captions

a. Countries of recruitment

b. Dilemmas



c. Voting task



920

Figure 1. ***Overview of Experimental Methods***. *(A) Countries of recruitment for online samples nationally representative with respect to age and gender. (B) Summary of the five COVID-19 dilemmas employed in the experimental tasks. (C) Voting task: participants are asked to vote for a leader who will be entrusted with a group's charitable donation and can "embezzle" some of the donation money for themselves.*

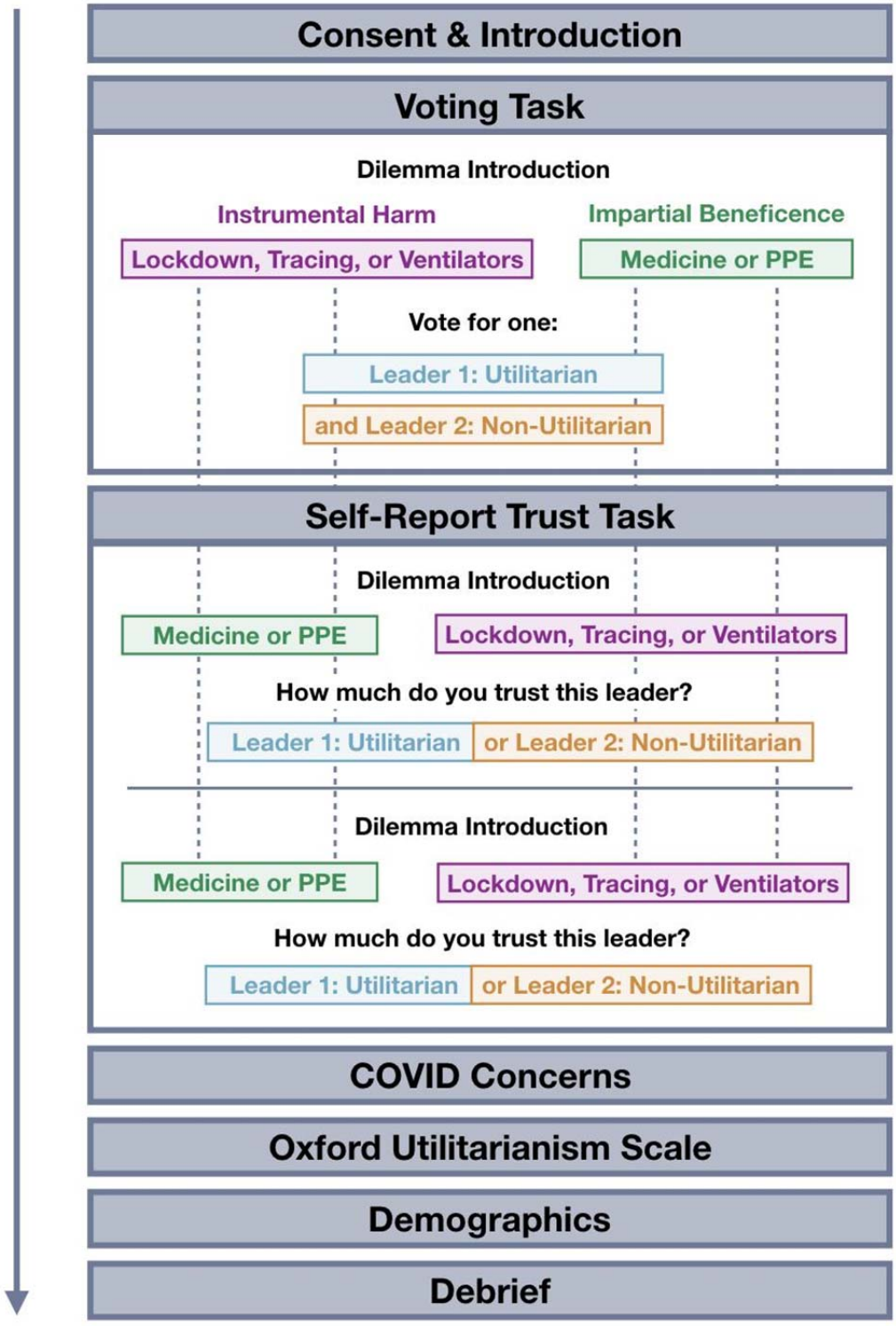


Figure 2. Overview of experimental design*. Across subjects, we will randomize the order of the voting and self-report tasks, the order of dilemmas in the self-report task, and the order of leaders in the voting task.*

Tables

Table 1. Design Table

Question	Hypothesis	Sampling Plan	Analysis Plan	Interpretation
How do leaders' endorsements of utilitarian vs. non-utilitarian approaches to dilemmas affect self-reported trust?	Participants will report higher trust in leaders who reject (versus endorse) instrumental harm, and higher trust in leaders who endorse (versus reject) impartial beneficence.	Power analyses suggested that our planned sample size (N = 12600) would be sufficient to achieve 95% power to detect an effect size of $d = .05$.	We will conduct a linear mixed-effects model of the effect of argument type, dimension type, and their interaction on the composite measure of self-report trust, controlling for demographic variables and participants' own policy preferences.	A significant interaction (after following up with post-hoc tests) will be interpreted as evidence that endorsement of instrumental harm decreases general trust in leaders, while endorsement of impartial beneficence increases general trust in leaders. Should the observed effect size of the interaction be statistically equivalent (with the larger of the two p values in the TOST using equivalence bounds set by the SESOI smaller than alpha = .05), this will be interpreted as evidence for null effects.
How do leaders' endorsements of utilitarian vs. non-utilitarian	Participants will be more likely to vote for a leader who	Power analyses suggested that our planned sample size (N	We will conduct a generalized linear mixed-effects model	A significant main effect of dimension will be interpreted as

approaches to dilemmas affect trusting behavior?	rejects (versus endorses) instrumental harm, while they will be more likely to vote for a leader who endorses (versus rejects) impartial beneficence.	= 12600) would be sufficient to achieve 95% power to detect an odds ratio (*OR)* of 1.30.	with the logit link of the effect of dimension type on leader choice, controlling for demographic variables and participants' own policy preferences.	evidence that endorsement of instrumental harm decreases trusting behavior toward leaders, while endorsement of impartial beneficence increases trusting behavior toward leaders. Should the observed effect size be statistically equivalent (with the larger of the two p values in the TOST using equivalence bounds set by the SESOI smaller than alpha = .05), this will be interpreted as evidence for null effects.

935
936
937
938 **Table 2. Summary of moral arguments in COVID-19 dilemmas**
939

Dilemma	**Argument Type**	
<u>Instrumental Harm (IH)</u>	**Utilitarian**	**Non-Utilitarian**
Lockdown	"We need to think about all the consequences. Preventing deaths from	"As leaders, our primary duty is to protect our citizens. We must think of

	COVID isn't all that matters, and continuing these prolonged restrictions will have a far worse effect on our overall well-being."	our responsibilities to one another, and we cannot sacrifice some of our most vulnerable people in pursuit of the greater good."
Ventilators	"We have to think about how we can do the most good with the resources we have, and that means prioritizing those people who have the best chance of recovering and living a long and healthy life."	"It's not our place to choose who lives. Everyone has the same right to receive equal access to treatment, and we cannot abandon our most vulnerable in an effort to save more lives."
Tracing	"We need to control the pandemic, and sometimes you have to sacrifice the right to privacy for the greater good."	"Everyone has a right to privacy, and we cannot sacrifice this right in an effort to control the pandemic."
Impartial Beneficence (IB)	**Utilitarian**	**Non-Utilitarian**
Medicine	"COVID-19 is a global pandemic that affects all humans equally. We need to be impartial and send treatment where it can achieve the greatest good."	"We have a right to use our own resources to help our own citizens before everyone else. Other countries can produce their own treatments for COVID-19."
Personal Protective Equipment	"COVID-19 is a global pandemic that affects all humans. We need to be impartial in how we distribute resources like PPE and send it where it can achieve the greatest good."	"We have a duty to protect our own citizens first, not everyone in the world. Other countries are responsible for protecting their own citizens from COVID-19."

Supplementary Information

Pilot Data

We conducted two pilots to establish both proof of concept and the feasibility of our methods (see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c for data and code).

Pilot 1. We recruited a convenience sample of British participants (N = 100, after exclusions N = 98; N females = 56, Mean age = 31.87) on Prolific in July 2020 in which participants read five different COVID moral dilemmas. In a within-subjects design, participants completed the "dilemma introduction" (see Methods) for each of five dilemmas presented in randomized order. Following the introduction to each dilemma, participants were presented with two leaders in randomized order: a "utilitarian" leader, who argued for a utilitarian policy solution to the dilemma, and a "non-utilitarian" leader who rejected it. For each leader, participants provided two ratings of trustworthiness ("How trustworthy do you think this person is?" and "How likely would you be to trust this person's advice on other issues?", both on a 7-point scale), which we averaged (separately for each participant, dilemma, and leader) to create a composite trust measure.

We conducted a linear mixed-effects model of the effect of argument type (Utilitarian vs. Non-Utilitarian), dimension type (Instrumental Harm vs. Impartial Beneficence), and their interaction, on the composite score of trust, adding demographic variables (namely race, gender, age, education level, income, political ideology, and religiosity), and policy support as fixed effects, and dilemmas and participants as random intercepts. For the purposes of the analysis, we used effect coding such that for argument type, the Non-Utilitarian condition is coded as -0.5 and the Utilitarian condition as 0.5, and for the dimension type, Instrumental Harm is coded as -0.5, and Impartial Beneficence as 0.5. The demographic covariates were grand mean-centered; the gender variable was dummy coded with "female" as baseline, and the race variable was dummy coded with "other" as baseline. P-values were computed using Satterthwaite's approximation for degrees of freedom as implemented in lmerTest. For analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

We observed no significant main effect of either argument type (B = -0.01, standard error [SE] = 0.08, $t(875)$ = -0.15, p = .881, confidence interval [CI] = [-0.17, 0.14]) or dimension type (B = -0.01, SE = 0.17, $t(3)$ = -0.07, p = .946, CI = [-0.34, 0.31]), but crucially, a significant interaction between argument and dimension type (B = 2.33, SE = 0.16, $t(875)$ = 14.67, $p < .001$, CI = [2.02, 2.64]). Post-hoc comparisons with Bonferroni corrections confirmed that in Instrumental Harm dilemmas, utilitarian leaders were seen as less trustworthy than non-utilitarian leaders (mean trust for utilitarian leaders = 3.48, SE = 0.24, CI = [2.86, 4.11]; mean trust for non-utilitarian leaders = 4.66, SE = 0.24, CI = [4.04, 5.28]; B = -1.18, SE = 0.10, $t(875)$ = -11.72, $p < .001$, CI = [-1.37, -0.98]), but in Impartial Beneficence dilemmas this effect was reversed, such that utilitarian leaders were seen as more trustworthy than non-utilitarian leaders

(mean trust for utilitarian leaders = 4.64, SE = 0.25, CI = [3.97, 5.30]; mean trust for non-utilitarian leaders = 3.48, SE = 0.25, CI = [2.82, 4.15]; B = 1.15, SE = 0.12, $t(875) = 9.37$, $p <$ .001, CI = [0.91, 1.39]; see Supplementary Figure 2; for results by dilemma, see Supplementary Figure 3).

Pilot 2. For the second pilot experiment, which included both the voting task and the self-report trust task, we recruited a convenience sample of U.S. participants (N = 503, after exclusions N = 469; N females = 239, Mean age = 30.33) on Prolific in July 2020. They completed a procedure mostly identical to that described in the Methods section, with the exception that this pilot did not include attention or comprehension checks, it did not include the Tracing dilemma, and we used The Red Cross instead of UNICEF in the voting task. Participants in this pilot experiment were voter participants in the voting task. A few days prior to running the main pilot experiment, we recruited a convenience sample of donor participants (total N = 103, after exclusions N = 100; N females = 59, Mean age = 30.16) via Prolific. The donor participants chose to contribute a total of $71.80 to The Red Cross. We displayed this amount to voter participants in the main pilot experiment.

Following the analysis plan we will employ in the Registered Report (see Analysis Plan), for the self-reported trust task, we conducted a linear mixed-effects model of the effect of argument type (Utilitarian vs. Non-Utilitarian), dimension type (Instrumental Harm vs. Impartial Beneficence), and their interaction, on the composite score of trust, adding demographic variables (namely race, gender, age, education level, income, political beliefs, and religiosity) and policy support as fixed effects, and dilemmas and participants as random intercepts. For the purposes of the analysis, we used effect coding such that for argument type, the Non-Utilitarian condition was coded as -0.5 and the Utilitarian condition as 0.5, and for the dimension type, Instrumental Harm is coded as -0.5, and Impartial Beneficence as 0.5. The demographic covariates were grand mean-centered; the gender variable was dummy coded with "female" as baseline, and the race variable was dummy coded with "other" as baseline. P-values were computed using Satterthwaite's approximation for degrees of freedom as implemented in lmerTest. For analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

For self-reported trust, there was a significant main effect of argument type (B = -0.58, SE = 0.12, $t(452) = -4.75$, $p <$ .001, CI = [-0.82, -0.35]), no main effect of dimension type (B = 0.10, SE = 0.27, $t(3) = 0.38$, p = .730, CI = [-0.41, 0.62]), and crucially, a significant interaction between argument and dimension type (B = 2.88, SE = 0.24, $t(452) = 11.80$, $p <$ .001, CI = [2.41, 3.35]). Post-hoc comparisons with Bonferroni corrections confirmed that in Instrumental Harm dilemmas, utilitarian leaders were seen as less trustworthy than non-utilitarian leaders (mean trust for utilitarian leaders = 3.14, SE = 0.24, CI = [2.33, 3.95]; mean trust for non-utilitarian leaders = 5.16, SE = 0.24, CI = [4.35, 5.96]; B = -2.02, SE = 0.17, $t(454) = -11.59$, $p <$ .001, CI = [-2.36, -1.68]), but in Impartial Beneficence dilemmas this effect was reversed, such that utilitarian leaders were seen as more trustworthy than non-utilitarian leaders (mean trust for

utilitarian leaders = 4.68, SE = 0.24, CI = [3.88, 5.48]; mean trust for non-utilitarian leaders = 3.82, SE = 0.24, CI = [3.02, 4.63]; $B = 0.86$, SE = 0.17, $t(455) = 5.00$, $p < .001$, CI = [0.52, 1.19]; see Supplementary Figure 4; for results by dilemma, see Supplementary Figure 5).

For the voting task, we first excluded participants who reported not understanding the task (N = 17; remaining N = 452). Following the analysis plan we will employ in the Registered Report (see Analysis Plan) for the voting task, we ran a generalized linear mixed-effects model with the logit link of the effect of dimension type (Instrumental Harm vs. Impartial Beneficence) on the leader choice (Utilitarian vs. Non-Utilitarian), adding demographic variables (namely race, gender, age, education level, income, political beliefs, and religiosity) and policy support as fixed effects, and dilemma as a random intercept. For the purposes of the analysis, we used effect coding such that for the binary response variable of argument type, the Non-Utilitarian trust response is coded as 0 and the Utilitarian trust response as 1, and for the dimension type, Instrumental Harm is coded as -0.5, and Impartial Beneficence as 0.5. The demographic covariates were grand mean-centered; the gender variable was dummy coded with "female" as baseline, and the race variable was dummy coded with "other" as baseline. P-values were computed using Satterthwaite's approximation for degrees of freedom as implemented in lmerTest. For analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c.

We found a significant main effect for dimension type ($B = 2.41$, SE = 0.33, $z = 7.30$, $p < .001$, CI = [1.77, 3.13], OR = 11.13). Post-hoc comparisons with Bonferroni corrections confirmed that in Instrumental Harm dilemmas, participants were less likely to vote for utilitarian leaders than non-utilitarian leaders (probability of choosing utilitarian leader = 0.15, SE = 0.05, CI = [0.06, 0.31]), but in Impartial Beneficence dilemmas this effect was reversed, such that participants were more likely to vote for utilitarian leaders than non-utilitarian leaders (probability of choosing utilitarian leader = 0.65, SE = 0.08, CI = [0.46, 0.81]; see Supplementary Figure 6). In other words, participants were more than 11 times more likely to choose the utilitarian leader in Impartial Beneficence dilemmas compared to Instrumental Harm dilemmas.

We note here that this model yielded a singular fit, due to the addition of dilemmas as a random intercept. First, we confirmed that a more parsimonious model, identical in every way except for the omission of the random intercept, yielded convergent results (for analysis code, see https://osf.io/m9tpu/?view_only=2f5896f4b1c14be687db03d33083976c). Given the theoretical importance of including dilemmas as a random intercept, we report here the results of the more theoretically appropriate maximal random effects structure, which should be preferred when justified by the design[72].

Supplementary Methods

1. Supplementary Method: Dilemma Preambles

Lockdown Dilemma

Dilemma Preamble

Think ahead several months into the future. Imagine that the U.S. is in the middle of another wave of the COVID-19 pandemic, and there is still no vaccine available. After a surge in cases, political leaders have imposed strict nationwide restrictions, with stay-at-home orders and closures of schools, offices, shops, restaurants, bars, theaters, and so on.

These restrictions have now been in place for three months. It's clear that the policy is working to reduce the number of deaths, especially amongst vulnerable people. However, these prolonged restrictions are taking their toll. Mental health experts and economists are increasingly concerned about the effects of continued restrictions on people's overall wellbeing. Because of this, political leaders are debating when to lift the restrictions and reopen schools and businesses.

Some are arguing that we should consider lifting the restrictions immediately. They argue that even though resuming activities now will cause more COVID-related deaths in the short-term, the economic and social consequences of continuing the prolonged restrictions could cause worse suffering overall in the long term.

Others are arguing that the restrictions should stay in place at least until a vaccine is available. They argue that the country has a primary responsibility to protect its vulnerable citizens, and that this must take priority.

Support Measure

Which policy do you think should be adopted?

1	2	3	4	5	6	7
Strongly support keeping the prolonged restrictions			Indifferent			Strongly support lifting the restrictions

Morality Measure (exploratory)

How morally right or wrong would it be to lift restrictions before a vaccine is available?

1	2	3	4	5	6	7
Absolutely morally wrong			Neither right nor wrong			Absolutely morally right

Ventilators Dilemma

Dilemma Preamble

COVID-19 remains a public health threat. Public health officials have announced that citizens should be on alert for another dangerous wave of the pandemic.

If the predictions are correct, there will not be enough ventilators and hospital beds to treat everyone, and doctors are going to have to make difficult decisions about how to ration medical care. Political leaders are calling for a policy to be put into place now so the same standards can be applied in hospitals across the country.

Some are arguing that when allocating access to ventilators and other forms of healthcare, doctors should prioritize younger and healthier people because they are more likely to survive treatment.

Others are arguing that everyone should have equal access to treatment, regardless of their age or health status.

Support Measure

Which policy do you think should be adopted?

1	2	3	4	5	6	7
Strongly support everyone having equal access to treatment			Indifferent			Strongly support prioritizing younger and healthier people for treatment

Morality Measure (exploratory)

How morally right or wrong would it be to prioritize younger and healthier people for COVID treatment?

1	2	3	4	5	6	7
Absolutely morally wrong			Neither right nor wrong			Absolutely morally right

Tracing Dilemma

Dilemma Preamble

COVID-19 remains a threat to public health. Scientists are suggesting that an effective way to prevent the spread of COVID-19 is through wide-reaching and mandatory "contact tracing." This works by using technology to trace people's movements and interactions with other people. If someone tests positive for COVID-19, the technology can alert others who have been in contact with that person.

Public officials are considering a new contact tracing program that goes much further than current contact tracing efforts. This new program is estimated to be more effective at containing the pandemic, but is also more invasive of individual privacy. This proposal involves delivering inexpensive contact tracing devices to each resident. The small devices, which don't require a mobile phone and can be worn on a lanyard or carried in a handbag, use GPS and cellular technology to continuously trace the wearer's movements. The new program would require residents to carry a tracing device whenever they leave their homes, and residents could be fined if they fail to bring the device with them.

Some are arguing that the government should make it mandatory for individuals to carry tracing devices with them whenever they leave their homes. They are saying that sometimes you have to sacrifice privacy for the greater good.

Others argue that these tracing devices should be only voluntary, because forcing residents to wear them anytime they leave their homes would violate their rights to privacy.

Support Measure

Which policy do you think should be adopted?

1	2	3	4	5	6	7
Strongly support the tracing devices being voluntary			Indifferent			Strongly support the tracing devices being mandatory

Morality Measure (exploratory)

How morally right or wrong would it be to make it mandatory for individuals to carry contact tracing devices with them wherever they go?

1	2	3	4	5	6	7
Absolutely morally wrong			Neither right nor wrong			Absolutely morally right

Medicine Dilemma

Dilemma Preamble

COVID-19 remains a public health threat. Public health officials have announced that citizens should be on alert for another dangerous wave of the pandemic.

Imagine that a pharmaceutical company based in the U.S. has developed an effective treatment. The company is manufacturing the medicine as quickly as possible, but it is unlikely there will be sufficient supplies when the next wave hits. Political leaders are debating how the medicine should be distributed around the globe.

Some are arguing that the medicine should be sent wherever it can achieve the greatest good, even if that means sending it to other countries.

Others are arguing that the medicine should be kept in the U.S., because the government should focus on protecting its own citizens.

Support Measure

Which policy do you think should be adopted?

1	2	3	4	5	6	7
Strongly support U.S.-made medicine being reserved for protecting American citizens			Indifferent			Strongly support U.S.-made medicine being given to whoever needs it most

Morality Measure (exploratory)

How morally right or wrong would it be for U.S.-made medicine to be given to whoever needs it most, even if that means sending it to other countries?

1	2	3	4	5	6	7

Absolutely morally wrong	Neither right nor wrong	Absolutely morally right

PPE Dilemma

Dilemma Preamble

Think ahead several months into the future. Imagine that COVID-19 cases are rising again around the world and public health officials have announced that citizens should be on alert for another dangerous wave of the pandemic.

If the predictions are correct, there will soon be another global shortage of personal protective equipment (PPE). Political leaders are debating how personal protective equipment should be distributed around the globe.

Some are arguing that PPE made in American factories should be sent wherever it can do the most good, even if that means sending it to other countries.

Others are arguing that PPE made in American factories should be kept in the U.S., because the government should focus on protecting its own citizens.

Support Measure

Which policy do you think should be adopted?

1	2	3	4	5	6	7
Strongly support U.S.-made PPE being reserved for protecting American citizens			Indifferent			Strongly support U.S.-made PPE being given to whoever needs most

Morality Measure (exploratory)

How morally right or wrong would it be for U.S.-made PPE to be given to whoever needs it most, even if that means sending it to other countries?

1	2	3	4	5	6	7
Absolutely morally wrong			Neither right			Absolutely morally right

nor wrong

1225

2. Supplementary Method: Voting Task

Voter Participants

Note: Order of utilitarian and non-utilitarian leaders will be randomized (i.e., randomly assigned to Person 1 and Person 2 in the instructions below).

Introduction (displayed for all dilemmas)

We now ask you to make a choice that has real financial consequences. A few days ago, a group of 100 people were recruited via an international online marketplace and invited to make donations to the charitable organization UNICEF. In total, they donated an amount equivalent to [AMOUNT].

We would like you to vote for a leader to be responsible for the entire group's donations. The leader has two options:

- They can transfer the group's [AMOUNT] donation to UNICEF in full, *or*
- They can take some of this money for themselves (up to the full amount) and transfer whatever amount is left to UNICEF.

Lockdown Dilemma

We now want you to vote between two people to be the leader responsible for the donations. We showed both people the debate you just read about reopening schools and businesses, and asked them what they thought.

1. **Person 1 believes that restrictions should stay in place at least until a vaccine is available.** They agreed with the statement, "*As leaders, our primary duty is to protect our citizens. We must think of our responsibilities to one another, and we cannot sacrifice some of our most vulnerable people in pursuit of the greater good*."
2. **Person 2 believes that we should consider lifting the restrictions immediately.** They agreed with the statement, "*We need to think about all the consequences. Preventing deaths from COVID isn't all that matters, and continuing these prolonged restrictions will have a far worse effect on our overall well-being.*"

Please vote for which person you wish to be the leader. For the person who gets the majority number of votes, we will implement their decision and make the payments accordingly.

- Person 1, the person who believes restrictions should stay in place at least until a vaccine is available.
- Person 2, the person who believes we should consider lifting the restrictions immediately.

Ventilators Dilemma

We now want you to vote between two people to be the leader responsible for the donations. We showed both people the debate you just read about how to distribute healthcare resources, and asked them what they thought.

1. **Person 1 believes that younger and healthier people should be prioritized for COVID treatment.** They agreed with the statement, *"We have to think about how we can do the most good with the resources we have, and that means prioritizing those people who have the best chance of recovering and living a long and healthy life."*
2. **Person 2 believes that doctors should give everyone equal access to COVID treatment.** They agreed with the statement, *"It's not our place to choose who lives. Everyone has the same right to receive equal access to treatment, and we cannot abandon our most vulnerable in an effort to save more lives."*

Please vote for which person you wish to be the leader. For the person who gets the majority number of votes, we will implement their decision and make the payments accordingly.

- Person 1, the person who believes that younger and healthier people should be prioritized for COVID treatment.
- Person 2, the person who believes that doctors should give everyone equal access to COVID treatment.

Tracing Dilemma

We now want you to vote between two people to be the leader responsible for the donations. We showed both people the debate you just read about mandatory contact tracing devices, and asked them what they thought.

1. **Person 1 believes that it should be mandatory for residents to carry tracing devices whenever they leave their homes.** They agreed with the statement, *"We need to control the pandemic, and sometimes you have to sacrifice the right to privacy for the greater good."*
2. **Person 2 believes that it should be voluntary for residents to carry tracing devices whenever they leave their homes.** They agreed with the statement, *"Everyone has a right to privacy, and we cannot sacrifice this right in an effort to control the pandemic."*

Please vote for which person you wish to be the leader. For the person who gets the majority number of votes, we will implement their decision and make the payments accordingly.

- Person 1, the person who believes that it should be mandatory for residents to carry tracing devices whenever they leave their homes.
- Person 2, the person who believes that it should be voluntary for residents to carry tracing devices whenever they leave their homes.

Medicine Dilemma

We now want you to vote between two people to be the leader responsible for the donations. We showed both people the debate you just read about distributing medicine for COVID, and asked them what they thought.

1. **Person 1 believes that U.S.-made medicine should be reserved for treating American citizens.** They agreed with the statement, "*We have a right to use our own resources to help our own citizens before everyone else. Other countries can produce their own treatments for COVID-19.*"
2. **Person 2 believes that U.S.-made medicine should be given to whoever needs it most, even if that means sending it to other countries.** They agreed with the statement, "*COVID-19 is a global pandemic that affects all humans equally. We need to be impartial and send treatment where it can achieve the greatest good."*

Please vote for which person you wish to be the leader. For the person who gets the majority number of votes, we will implement their decision and make the payments accordingly.

- Person 1, the person who believes that U.S.-made medicine should be reserved for treating American citizens.
- Person 2, the person who believes that U.S.-made medicine should be given to whoever needs it most, even if that means sending it to other countries.

PPE Dilemma

We now want you to vote between two people to be the leader responsible for the donations. We showed both people the debate you just read about keeping personal protective equipment in the U.S., and asked them what they thought.

1. **Person 1 believes that U.S.-made PPE should be given to whoever needs it most, even if that means sending it to other countries.** They agreed with the statement, *"COVID-19 is a global pandemic that affects all humans. We need to be impartial in how we distribute resources like PPE and send it where it can achieve the greatest good."*
2. **Person 2 believes that U.S.-made PPE should be reserved for protecting American citizens.** They agreed with the statement, *"We have a duty to protect our own citizens*

first, not everyone in the world. Other countries are responsible for protecting their own citizens from COVID-19."

Please vote for which person you wish to be the leader. For the person who gets the majority number of votes, we will implement their decision and make the payments accordingly.

- Person 1, the person who believes that U.S.-made PPE should be given to whoever needs it most, even if that means sending it to other countries.
- Person 2, the person who believes that U.S.-made PPE should be reserved for protecting American citizens.

Donor Participants

Donation Task

We're giving you a choice to allocate some real money. We are awarding you a $2.00 bonus on top of your payment for participating in this study. You have the choice of how much of this bonus you want to keep for yourself, and how much you'd like to donate to **UNICEF**.

UNICEF is a humanitarian organization working on the ground in more than 190 countries, partnering with front-line responders and providing them with the information and resources they need to keep children healthy and learning and protected from sickness and violence during the pandemic.

How much of your $2.00 bonus would you like to donate to UNICEF?

Whatever is remaining will be added to your total payment.

I would like to donate...

0 100

Embezzlement Task

Earlier in this study, we gave you the opportunity to donate to UNICEF. We are recruiting 100 participants in this study, who all have the chance to make donations. After we get all the donation decisions, we are going to select one participant to be responsible for the donations of the whole group.

Because each of you can choose to donate an amount between $0 and $2, the total amount you will be responsible for if you are selected to be responsible for the group will range from $0 to $200.

If you are selected to be responsible for the group, you can choose to transfer 100% of the donation money to UNICEF. Or, you can choose to keep some of the money for yourself, and transfer the rest to UNICEF. Any money you choose to keep from the group's total would be added to your bonus.

If you are selected to be responsible for the group's donations, what percentage of the total donations do you want to keep as an additional bonus?

I would like to **keep** _% of the total amount

0 100

3. Supplementary Method: Self-Report Trust Task

Note: Selection of either the utilitarian or non-utilitarian leader will be randomized. The wording and response options for the questions on trust are the same across dilemmas.

Lockdown Dilemma

Non-Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that restrictions should stay in place at least until a vaccine is available.

This mayor said, "*As leaders, our primary duty is to protect our citizens. We must think of our responsibilities to one another, and we cannot sacrifice some of our most vulnerable people in pursuit of the greater good.*"

How trustworthy do you think this person is?

1	2	3	4	5	6	7
Not at all trustworthy			Somewhat trustworthy			Extremely trustworthy

How likely would you be to trust this person's advice on other issues?

1	2	3	4	5	6	7
Not at all likely			Somewhat likely			Extremely likely

Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that we should consider lifting the restrictions immediately.

This mayor said, "*We need to think about all the consequences. Preventing deaths from COVID isn't all that matters, and continuing these prolonged restrictions will have a far worse effect on our overall well-being.*"

How trustworthy do you think this person is?

1	2	3	4	5	6	7
Not at all trustworthy			Somewhat trustworthy			Extremely trustworthy

How likely would you be to trust this person's advice on other issues?

1	2	3	4	5	6	7
Not at all likely			Somewhat likely			Extremely likely

Ventilators Dilemma

Non-Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that doctors should give everyone equal access to COVID treatment.

This mayor said, *"It's not our place to choose who lives. Everyone has the same right to receive equal access to treatment, and we cannot abandon our most vulnerable in an effort to save more lives."*

Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that younger and healthier people should be prioritized for COVID treatment.

This mayor said, *"We have to think about how we can do the most good with the resources we have, and that means prioritizing those people who have the best chance of recovering and living a long and healthy life."*

Tracing Dilemma

Non-Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that it should be voluntary for residents to carry contact tracing devices whenever they leave their homes.

This mayor said, *"Everyone has a right to privacy, and we cannot sacrifice this right in an effort to control the pandemic."*

Utilitarian Leader

Imagine that the mayor of a major city in your area is arguing that it should be mandatory for residents to carry contact tracing devices whenever they leave their homes.

This mayor said, *"We need to control the pandemic, and sometimes you have to sacrifice the right to privacy for the greater good."*

Medicine Dilemma

Non-Utilitarian Leader

Imagine that the mayor of a major city in your region is arguing that U.S.-made medicine should be reserved for protecting American citizens.

This mayor said, *"We have a right to use our own resources to help our own citizens before everyone else. Other countries can produce their own treatments for COVID-19."*

Utilitarian Leader

Imagine that the mayor of a major city in your region is arguing that U.S.-made medicine should be given to whoever needs it most, even if that means sending it to other countries.

This mayor said, *"COVID-19 is a global pandemic that affects all humans equally. We need to be impartial and send treatment where it can achieve the greatest good."*

PPE Dilemma

Non-Utilitarian Leader

Imagine that the mayor of a major city in your region is arguing that U.S.-made PPE should be reserved for protecting American citizens.

This mayor said, *"We have a duty to protect our own citizens first, not everyone in the world. Other countries are responsible for protecting their own citizens from COVID-19."*

Utilitarian Leader

Imagine that the mayor of a major city in your region is arguing that U.S.-made PPE should be given to whoever needs it most, even if that means sending it to other countries.

This mayor said, *"COVID-19 is a global pandemic that affects all humans. We need to be impartial in how we distribute resources like PPE and send it where it can achieve the greatest good."*

Supplementary Notes

1. **Utilitarian versus non-utilitarian approaches to moral dilemmas.**
2. **Utilitarian arguments in the lockdown dilemma.**
3. **How this work advances understanding of moral dilemmas and trust.**
4. **On the intuitiveness of our hypotheses.**
5. **Potential cross-cultural differences in our study.**
6. **Impartiality in the Ventilators, Medicine and PPE dilemmas.**
7. **Ruling out a general preference for leaders who are less restrictive.**
8. **Ruling out a general preference for leaders who treat everyone equally.**
9. **Ruling out a general preference for leaders who seek to minimize COVID-related deaths.**
10. **Generalizability of trust measures.**
11. **The relationship between the voting task and the concept of impartial beneficence.**
12. **Why we chose UNICEF as the charity in the voting task.**
13. **Ruling out a general preference for leaders who participants agree with on policy issues.**

1. Utilitarian versus non-utilitarian approaches to moral dilemmas.

In moral psychology, moral dilemmas are often characterised in terms of a conflict between utilitarianism and deontology. While deontological approaches are the most commonly discussed counterpoint to utilitarianism, there are other ethical approaches that are neither deontological nor utilitarian, such as virtue ethics. Moreover, there are many different types of deontological theories, and while these cohere in agreeing that there is more to morality than the utilitarian impartial maximization of welfare, they often disagree on the specific details. Therefore, for precision, and to avoid inviting conclusions about specific deontological approaches to morality, in this paper we refer to "utilitarian" and "non-utilitarian" agents.

2. Utilitarian arguments in the lockdown dilemma.

Aren't there good utilitarian arguments for lockdowns? Why does the utilitarian leader argue for lifting restrictions in your Lockdown dilemma?

Utilitarianism says that what matters is what brings about the best consequences, and wherever there is disagreement about what would have the best overall consequences, there can be disagreement on utilitarian grounds for what action is correct.

In the case of lockdown, it is certainly possible to construct both utilitarian and non-utilitarian/deontological arguments for pro- and anti-lockdown positions, particularly because there are many types of lockdowns that have been implemented around the world during this pandemic. For example, one could argue that we need to sacrifice individual freedoms for the greater good, to prevent hospitals from overflooding. Or one could argue that the economic consequences of a lockdown are less bad than the economic consequences of letting the pandemic run rampant.

In the current work, we test a very specific case of lockdown that maps more clearly onto utilitarian and non-utilitarian arguments. Our dilemma describes a lockdown that has been dragging on for three months, with no clear end in sight, that is having a clearly negative impact on citizens' wellbeing. (This specific dilemma resembles a situation in the Philippines, where citizens endured a continuous lockdown for more than three months and citizens there reported historic lows in wellbeing.)

In the specific dilemma that we are testing, the anti-lockdown position is clearly argued on utilitarian grounds and the pro-lockdown position is clearly argued on non-utilitarian, deontological grounds. We constructed these arguments based on public statements made by political leaders and prominent utilitarian philosophers like Peter Singer, who have quite consistently argued *against* lockdown using utilitarian arguments (e.g. "*It pains us to say it, but US President Donald Trump is right. We can't let the cure be worse than the disease. Lockdowns have health benefits: fewer will die of COVID-19, as well as other transmissible diseases. But they have real social and economic costs, [including] social isolation, unemployment, and widespread bankruptcies": Singer & Plant, April 6th 202*0).

Future studies might focus on exploring lockdown dilemmas more extensively by investigating how utilitarian and deontological arguments for and against lockdowns impact trust in leaders. In the current work, our goal was to investigate patterns of trust in leaders across a variety of pandemic dilemmas, rather than just focusing on one dilemma. If we see consistent mistrust in utilitarian leaders across Lockdown, Ventilators and Tracing dilemmas (as we saw in our pilot data), we can reasonably conclude that endorsement of instrumental harm reduces trust in leaders. These three dilemmas are very different, but what they have in common is a tension between instrumental harm and maximizing aggregate welfare.

3. How this work advances understanding of moral dilemmas and trust.

There is growing evidence that utilitarian responses to moral dilemmas impact trust (e.g. Bostyn & Roets, 2017; Everett, Faber, Savulescu, & Crockett, 2018; Everett, Pizarro, & Crockett, 2016; Rom, Weiss, & Conway, 2017; Sacco et al. 2017; Uhlmann, 2013). However, this past research has several limitations:

1. It has been conducted using highly artificial hypothetical moral dilemmas (such as the trolley problem) that most people will never encounter in their daily lives.
2. It has focused mainly on trust in anonymous strangers and largely ignored trust in political leaders.
3. It has focused mainly on how endorsement of instrumental harm in sacrificial dilemmas impacts trust, and has for the most part ignored the positive component of utilitarianism, impartial beneficence.
4. It has only been conducted in a limited number of Western populations (the US, UK and Germany).

The proposed work provides a significant advance by studying how both instrumental harm and impartial beneficence impact trust in leaders, by studying real-life dilemmas in the context of an ongoing global crisis, and by testing our hypothesis across a diverse set of populations around the globe. By grounding our work in prior theory and evidence, the present studies will be not only relevant to understanding human behavior in the current pandemic, but also in global crises more broadly, including future pandemics and climate change. Below we elaborate on these points.

Beyond artificial hypothetical dilemmas

Previous work on moral dilemmas and trust has mostly used artificial "trolley-style" moral dilemmas in which the target must make a decision about whether it's morally acceptable to save lives by, for example, pushing a large man off a footbridge to stop a runaway train or using lab assistants as human guinea pigs to see which of two mislabelled substances will kill. While such dilemmas are valuable tools in moral philosophy and psychology, most people will not have encountered dilemmas such as these in their daily lives and therefore the findings might not generalize to "real" moral dilemmas such as those that arise during a global health crisis. The COVID-19 pandemic has brought to bear numerous *real* moral dilemmas that are being debated by *real* political leaders, covered in *real* media outlets and followed by *real* people all around the world. These real dilemmas are the focus of the current work. By studying them, we can determine the extent to which past findings based on hypothetical, artificial dilemmas generalize to real dilemmas.

Advancing knowledge of trust in leaders.

We study how responses to moral dilemmas shape trust in leaders, moving beyond the previous focus on how ordinary people are evaluated based on their moral judgments. Most previous work has focused on trust in dyads, looking at how we infer the moral character of ordinary people who make decisions in moral dilemmas. Yet we know that utilitarianism differentially impacts perceptions of ordinary people and political leaders (Everett et al. 2018), which means we cannot generalize from past research on trust in utilitarians to a leadership context. Consider, for example, the (perhaps apocryphal) story of Winston Churchill who was told that the city of Coventry would be heavily bombed, and was faced with a decision to evacuate or not. If he evacuated the city the residents would be safe, but this might potentially reveal to the Germans that their code had been cracked. If he left the residents to their fate they would suffer great harm, but the secret of the code-breaking would remain intact and this, in turn, would likely lead to the war being over much sooner - saving many more thousands, if not millions, of lives. Churchill is said to have made the classic utilitarian calculation that it would be better to let some people suffer now for the greater good. That is, he endorsed instrumental harm - and is

celebrated as a national hero. Indeed, some work shows that utilitarians are perceived as more competent than non-utilitarians (Rom, Weiss, & Conway, 2017) and to the extent that trust in leaders is related to perceptions of their competence, it is possible that utilitarian approaches to pandemic dilemmas will *increase* rather than decrease trust in leaders. On balance however, the existing evidence suggests that political leaders who endorse instrumental harm would indeed be seen as *less* trustworthy - just like ordinary people who endorse instrumental harm are seen as less trustworthy and less suitable to be a political leader (Everett et al. 2018).

Beyond instrumental harm.

The vast majority of previous work on trust in utilitarians has focused on the negative dimension of utilitarianism (instrumental harm). But as outlined in the two-dimensional model of utilitarian psychology (Everett & Kahane, 2020; Kahane et al., 2018), utilitarianism involves more than just decisions about whether to sacrifice one to save a greater number. Instead, at the core of utilitarianism is the idea of *impartial beneficence*, that we must impartially maximise the well-being of all sentient beings on the planet in such a way that "[e]ach is to count for one and none for more than one" (Bentham, 1789/1983), not privileging compatriots, family members, or ourselves over strangers – or even enemies. Critically, these two dimensions of instrumental harm and impartial beneficence are both conceptually and psychologically distinct, with different psychological correlates (Kahane et al. 2015; Kahane et al. 2018) and there is evidence that they rely on different psychological processes (Capraro, Everett, & Earp, 2019). How would endorsement of impartial beneficence in moral dilemmas shape trust? Almost no work has considered this. The one exception is Everett et al (2018), who looked at perceptions of ordinary people who endorsed impartial beneficence (or instrumental harm) and found that impartial utilitarians were consistently *disfavored* for roles involving a direct interpersonal relationship, but that they were sometimes (but not always) preferred for distant, impersonal roles like a political leader. It is not clear, however, how robust this finding is, and whether we would see different results when looking at people explicitly described as political leaders - especially during a global crisis.

Generalizing across populations.

Past work on inferring trust from moral decisions has been conducted in just a handful of Western populations – the US, UK, and Germany – and so may not generalize to other countries that are also affected by the COVID-19 pandemic. Given well-publicized concerns about the WEIRDness of most published psychology research, it is important to move beyond such samples and establish the cultural generalizability of findings. Indeed, given observations of cultural variation in the willingness to endorse sacrificial harm, it is not a foregone conclusion that utilitarian decisions will impact trust in leaders universally.

4. On the intuitiveness of our hypotheses.

Is it not just obvious that people would trust leaders who reject instrumental harm? We think that there are good reasons to expect that utilitarian leaders who endorse instrumental harm would be trusted less based on previous empirical work and anecdotal data, but this is certainly not a foregone conclusion. Some work shows that people perceive those who endorse instrumental harm as less warm but more competent (Rom, Weiss, & Conway, 2017), and prefer

others who made characteristically utilitarian judgments for organizational leadership positions like a hospital manager (Rom, Weiss, & Conway, 2017), and other work shows that people strategically choose to endorse instrumental harm when the context favours competence-related traits, but are less likely to endorse instrumental harm when the context favours warmth-related traits (Rom & Conway, 2018). To the extent that political leadership requires competence, we might expect instead that people would favour leaders who make the decision to allow harm some to benefit the greater good - just as the wartime Prime Minister Winston Churchill is praised in the (probably apocryphal) story of allowing inhabitants of Coventry to be killed in order to shorten the war and thereby indirectly save many more lives (see Supplementary Note 3). Both possibilities are plausible, though on balance the existing evidence suggests that political leaders who endorse instrumental harm would indeed be seen as *less* trustworthy - just like ordinary people who endorse instrumental harm are seen as less trustworthy and less suitable to be a political leader (Everett et al. 2018).

Similarly, is it not just common sense that people would prefer leaders who endorse impartial beneficence? We don't think so. We predicted that people would trust leaders who endorse impartial beneficence more because people who endorse impartial beneficence are seen to make better political leaders, but not better friends, (Everett, Faber, Savulescu, & Crockett, 2018), consistent with other evidence that people do not endorse efficient maximization in charitable giving unless one is in a position of responsibility, like a political leader (Berman, Barasch, Levine, & Small, 2018). But we could also have predicted that leaders who endorse impartial beneficence would be trusted less precisely *because* they violate the norm of protecting their own citizens, since we know that intergroup partiality (here, "our" citizens vs. "others") is expected and favored in group leaders (e.g. Duck & Fielding, 1999, 2003). While overall both the existing evidence and anecdotal data suggest that impartial beneficence is more likely to increase trust, this is far from a foregone conclusion.

5. Potential cross-cultural differences in our study.

We hypothesize that endorsement of instrumental harm will reduce trust in leaders, while endorsement of impartial beneficence will increase trust in leaders. The main goal of our study is to assess the cross-cultural stability of this hypothesis. Given the required format for a Registered Report, we are limiting our focus to our specific, pre-registered predictions and not measuring other theoretical constructs that could potentially be relevant but which are outside our focus. We opted not to examine potential cross-cultural differences in this study because such differences might be confounded with pandemic severity, which differs dramatically across countries.

Nevertheless, it is certainly possible that there will be cross-cultural differences in our results. Should such differences emerge, we can conduct exploratory analyses with country-level indices (e.g. tightness/looseness, Human Development Index, pandemic severity, government policies related to the pandemic) but these are not discussed in the Stage 1 manuscript given the requirements of Registered Reports. We will make our data publicly available upon publication, so other researchers will be welcome to explore other questions with secondary analyses. Aside from testing our central question about trust in leaders, the data we're collecting will also be, to our knowledge, the largest cross-cultural dataset of moral judgments about pandemic dilemmas

and individual differences in utilitarianism, which we hope will be a valuable resource for other researchers.

6. Impartiality in the Ventilators, Medicine and PPE dilemmas.

Aren't the Ventilators, Medicine and PPE dilemmas all about impartiality? Why is the Ventilators dilemma considered an Instrumental Harm dilemma and not an Impartial Beneficence dilemma?

It's correct that all three of these dilemmas concern how to distribute resources to different people, and involve a tension between treating everyone equally versus prioritizing some people. However, there are key theoretical differences (which we also confirm empirically in our pilot data) between the Ventilators dilemma and the Medicine/PPE dilemmas that make them clear examples of instrumental harm and impartial beneficence, respectively.

The first crucial point is that impartial beneficence refers to a preference to impartially maximize aggregate welfare, and is therefore conceptually distinct from both generic prosociality and from non-maximizing impartiality. Allocating ventilators equally (i.e. regardless of personal characteristics) does not maximize aggregate welfare because older and sicker people are less likely to survive treatment, and have fewer years of quality life left to live. As Savulescu et al. (2020) describe in their paper applying utilitarian theory to the COVID pandemic, "*Utilitarianism would reject the idea of employing any form of 'first come, first served' to decide about treatment. The timing of when a patient arrives needing treatment is morally irrelevant to whether or not they should receive treatment… According to utilitarianism, doctors should be prepared to withdraw treatment from poor prognosis patients in order to enable the treatment of better prognosis patients if they arrive later*". Maximizing aggregate welfare is what matters from a utilitarian standpoint and, therefore, allocating ventilators equally is not a utilitarian policy, even though it is impartial. Moreover, prioritizing the young over the elderly is a utilitarian policy that involves instrumental harm: some people are denied treatment, or even have treatment taken away, in order to maximize aggregate welfare.

To validate empirically that the Ventilators dilemma does indeed tap instrumental harm and not impartial beneficence, we find that policy preferences in this dilemma correlate with the former and not the latter. In Pilot 2 we find that participants' own endorsement of prioritizing the young and healthy for Ventilators was significantly positively correlated with their instrumental harm score on the OUS ($r = 0.36$, $p < .001$) but not their impartial beneficence score ($r = -0.02$, $p = 0.737$).

In contrast, the Medicine and PPE dilemmas display the opposite pattern of results. Endorsement of sending resources where they are needed most was significantly positively correlated with OUS scores of impartial beneficence (Medicine: $r = 0.35$, $p < .001$; PPE: $r = 0.38$, $p < .001$), but not instrumental harm (Medicine: $r = -0.12$, $p = .022$; PPE: $r = 0.05$, $p = .390$).

For these reasons - both theoretical and empirical - we believe that the Ventilators dilemma is indeed tapping into the domain of instrumental harm rather than impartial beneficence, while the Medicine and PPE dilemmas tap impartial beneficence rather than instrumental harm.

7. Ruling out a general preference for leaders who are less restrictive.

Our theory predicts a very specific overall pattern of results across dilemmas: that people will trust the non-utilitarian leader more in Instrumental Harm dilemmas, while they will trust the utilitarian leader more in Impartial Beneficence dilemmas. It is also possible that people simply prefer leaders who are less restrictive. This might be especially relevant for democracies that place a strong priority on individual liberty and freedoms, such as the United States. Such a preference would predict that people will distrust leaders who impose lockdowns and mandatory contact tracing policies, and that people will distrust utilitarian and non-utilitarian leaders equally in the Ventilators, Medicine and PPE dilemmas. This overall pattern predicted by a preference for less restrictive leaders is not what we found in our pilots and it is not what we expect to see in our main study.

Dilemma	**Pilot results**	**Preference for less restrictive leaders**
Lockdown (IH)	Prefer non-utilitarian	Prefer utilitarian
Ventilators (IH)	Prefer non-utilitarian	No preference
Tracing (IH)	Prefer non-utilitarian	Prefer non-utilitarian
Medicine (IB)	Prefer utilitarian	No preference
PPE (IB)	Prefer utilitarian	Prefer utilitarian

8. Ruling out a general preference for leaders who treat everyone equally.

Our theory predicts a very specific overall pattern of results across dilemmas: that people will trust the non-utilitarian leader more in Instrumental Harm dilemmas, while they will trust the utilitarian leader more in Impartial Beneficence dilemmas. Another possibility is that people simply prefer leaders who treat everyone equally. Such a preference would predict that people will prefer leaders who allocate ventilators equally and distribute medicines and PPE impartially around the globe, and that people will distrust utilitarian and non-utilitarian leaders equally in the Lockdown and Tracing dilemmas. This overall pattern predicted by a preference for egalitarian leaders is not what we found in our pilots and it is not what we expect to see in our main study.

Dilemma	Pilot results	Preference for egalitarian leaders
Lockdown (IH)	Prefer non-utilitarian	No preference
Ventilators (IH)	Prefer non-utilitarian	Prefer non-utilitarian
Tracing (IH)	Prefer non-utilitarian	No preference
Medicine (IB)	Prefer utilitarian	Prefer utilitarian
PPE (IB)	Prefer utilitarian	Prefer utilitarian

9. Ruling out a general preference for leaders who seek to minimize COVID-related deaths.

Our theory predicts a very specific overall pattern of results across dilemmas: that people will trust the non-utilitarian leader more in Instrumental Harm dilemmas, while they will trust the utilitarian leader more in Impartial Beneficence dilemmas. It is also possible that people simply prefer leaders who seek to minimize deaths from COVID-19. Such a preference would predict a preference for leaders who impose lockdowns, prioritize the young over the elderly for ventilators, impose mandatory contact tracing, and distribute medicines and PPE impartially around the globe. This overall pattern predicted by a preference for leaders who seek to minimize COVID-related deaths is not seen in our pilots and we do not expect to find this in our main study.

Dilemma	Pilot results	Preference for leaders who minimize COVID deaths
Lockdown (IH)	Prefer non-utilitarian	Prefer non-utilitarian
Ventilators (IH)	Prefer non-utilitarian	Prefer utilitarian
Tracing (IH)	Prefer non-utilitarian	Prefer utilitarian
Medicine (IB)	Prefer utilitarian	Prefer utilitarian
PPE (IB)	Prefer utilitarian	Prefer utilitarian

10. Generalizability of trust measures.

Because our voting task involves asking participants to vote for a leader who will be responsible for a group donation, one might ask whether this measure can be generalized to overall trust in leaders, or if it is just specific to charity contexts.

Our voting task is not intended to measure trust in general, but this is by design. We designed this task to complement our self-reported measures of trust, which capture trust in general ("How trustworthy do you think this person is?", and "How likely would you be to trust this person's advice on other issues?").

We wanted to go beyond these self-report measures of general trust by including a behavioral measure of trust that involves real incentives. Any behavioral measure involving real incentives will necessarily involve a specific context; even the popular "Trust Game" involves the specific (and rather artificial) context of investing money with a stranger.

Because it is necessary to choose a specific context for a behavioral task, we considered many possibilities when designing the voting task. We chose a context that is highly relevant to our central research question: trust in leaders during a public health crisis. In the current pandemic, effective leadership involves being a *responsible steward of public resources in order to help those in need*. Our voting task measures how much people will trust someone to be *a responsible steward of a group's donations to help those in need*. We therefore think that the context we chose for our behavioral task bears directly on our research question and measures preferences for a specific type of leadership with clear relevance to the pandemic. Our pilot results suggest that these two types of measures (self-report and behavioural) tap a common core, with identical patterns in the predicted direction for all measures of trust, suggesting our results will generalize across diverse measures of trust.

11. The relationship between the voting task and the concept of impartial beneficence.

Because our voting task involves voting for a leader to be a responsible steward of a group's donations to help those in need, one might ask whether this task is too closely connected to the concept of impartial beneficence. The concept of impartial beneficence taps the endorsement of the impartial maximization of the greater good, even at the cost of personal self-sacrifice, and one example item in the impartial beneficence sub-scale of the Oxford Utilitarianism Scale is "It is morally wrong to keep money that one doesn't really need if one can donate it to causes that provide effective help to those who will benefit a great deal". It might be questioned whether our voting task, which involves voting for a leader who has the option to transfer a group's donations to a charity or to keep some money for themselves, might therefore be too closely connected to the construct of impartial beneficence.

It is important to note, however, that the participants in our main study do not themselves make any decisions about whether to donate to charity or not: the focus of our study is not charitable donation behavior. Rather, we ask whether people are more likely to trust utilitarian or non-utilitarian leaders to be responsible for other people's charitable donations -- a very different decision than a decision to donate to charity. Indeed, in our pilot results we find no evidence that participants' decisions in the voting task is influenced by their own endorsement of impartial beneficence: these scores did not predict choice of leader in the task, and the effects of leader argument on voting behavior remain significant when controlling for impartial beneficence.

12. Why we chose UNICEF as the charity in the voting task.

We put a lot of thought into our decision of which charity to use. Because we are running our study across 22 countries, we needed to select a charity that is internationally recognized and generally regarded as reliable and efficient across all countries in our sample. Not many charities fit these criteria; we originally selected The Red Cross/Crescent, but Pilot 2 results and comments from our collaborators in Asia suggested that this charity is seen as unreliable in many countries. We therefore settled on UNICEF as the best option.

13. Ruling out a general preference for leaders who participants agree with on policy issues.

We anticipated the possibility that people might simply prefer leaders who they agree with on policy issues. This is why, for each dilemma, we first ask participants which policy they prefer. We then control for individual policy support in all planned analyses. Our pilots using this analytic approach show that even after controlling for people's own policy preferences, the leader's policy argument impacts trust (see Pilot Data in Supplementary Information). What this means is that the leader's endorsement of instrumental harm or impartial beneficence in these dilemmas has a significant impact on trust, over and above the participant's own policy preference.

1866

1867
1868
1869
1870
1871

1872

1873

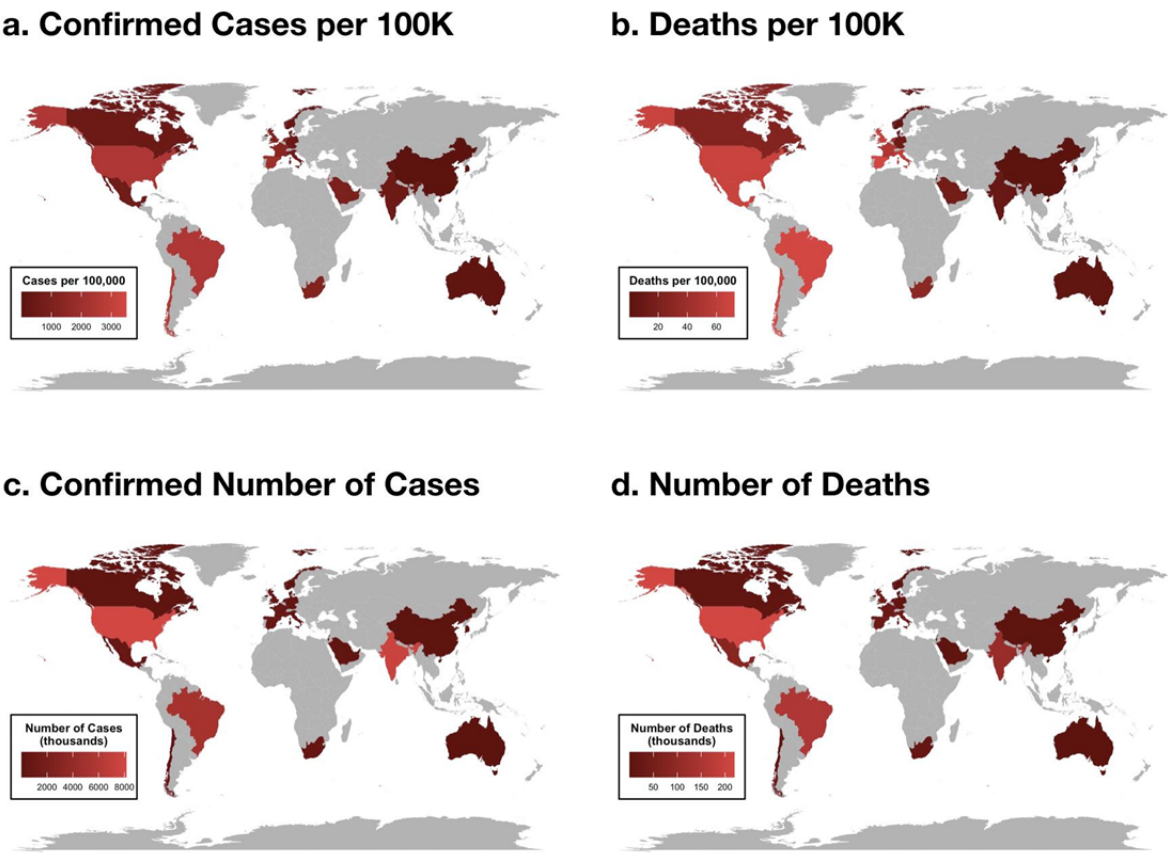
Supplementary Tables

Supplementary Table 1. Countries, recruitment platforms, survey languages, planned sample size, and planned sample characteristics. Participants in all countries will be able to select English as their language (in addition to the country's language, specified here in "Survey Language").

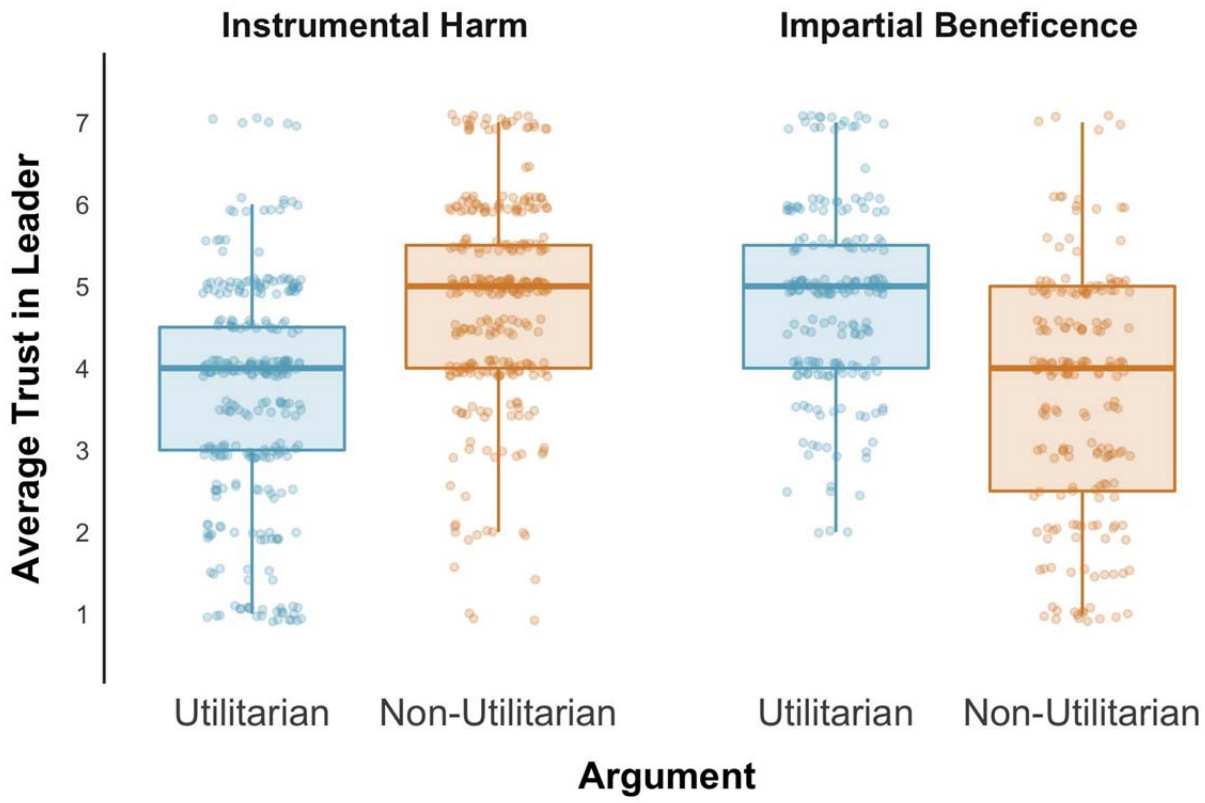
Country	Recruiting Platform	Survey Language	Planned N	Planned sample characteristics
Australia	Lucid	English	1000	Representative for age and gender
Brasil	Lucid	Portguese	1000	Representative for age and gender, expected shortfall age 50+
Canada	Lucid	English	1000	Representative for age and gender
Chile	CESS Santiago	Spanish	1000	Representative for age and gender
China	Lucid	Chinese	1000	Representative for gender; expected shortfall age 50+
Denmark	Epinion	Danish	1000	Representative for age and gender
France	Lucid	French	1000	Representative for age and gender
Germany	Lucid	German	1000	Representative for age and gender
India	Lucid	Hindi	1000	Representative for age and gender, expected shortfall age 50+
Israel	Panel HaMidgam	Hebrew	1000	Representative for age and gender
Italy	Lucid	Italian	1000	Representative for age and gender
Kingdom of Saudi Arabia	Lucid	Arabic	500	Expected 60% male, 90% age 18-44
Mexico	Lucid	Spanish	1000	Representative for age and gender, expected shortfall age 50+
Netherlands	Lucid	Dutch	1000	Representative for age and gender, expected shortfall age 50+
Norway	Norstat	Norweigan	1000	Representative for age and gender
Singapore	Lucid	English	1000	Representative for gender; expected shortfall age 50+
South Africa	Lucid	English	1000	Representative for gender; expected shortfall age 50+
South Korea	Lucid	Korean	1000	Representative for gender; expected shortfall age 50+
Spain	Lucid	Spanish	1000	Representative for age and gender

United Arab Emirates	Lucid	Arabic	500	Representative for gender; expected shortfall age 50+
United Kingdom	Lucid	English	1000	Representative for age and gender
United States of America	Lucid	English	1000	Representative for age and gender

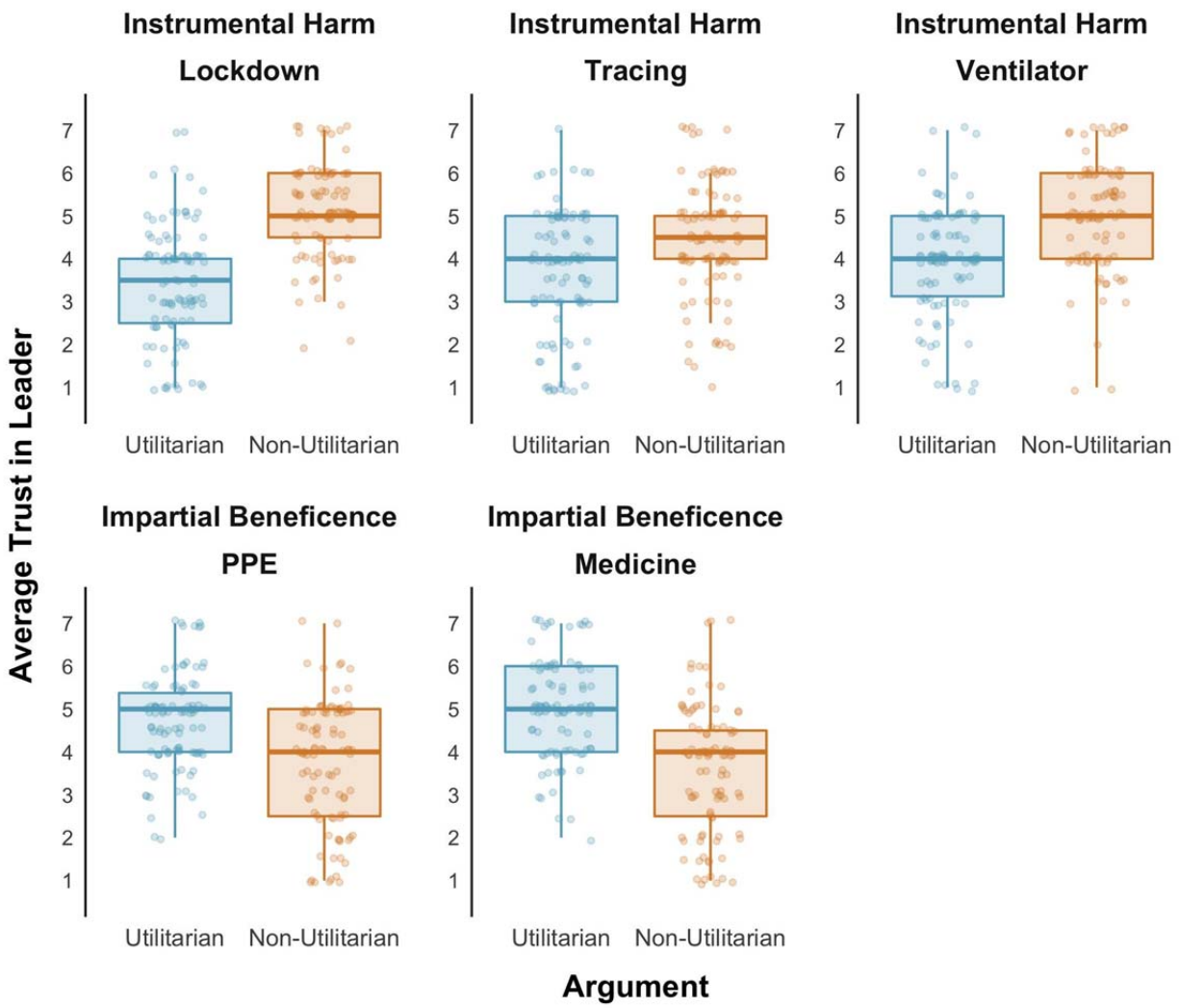
Supplementary Figures



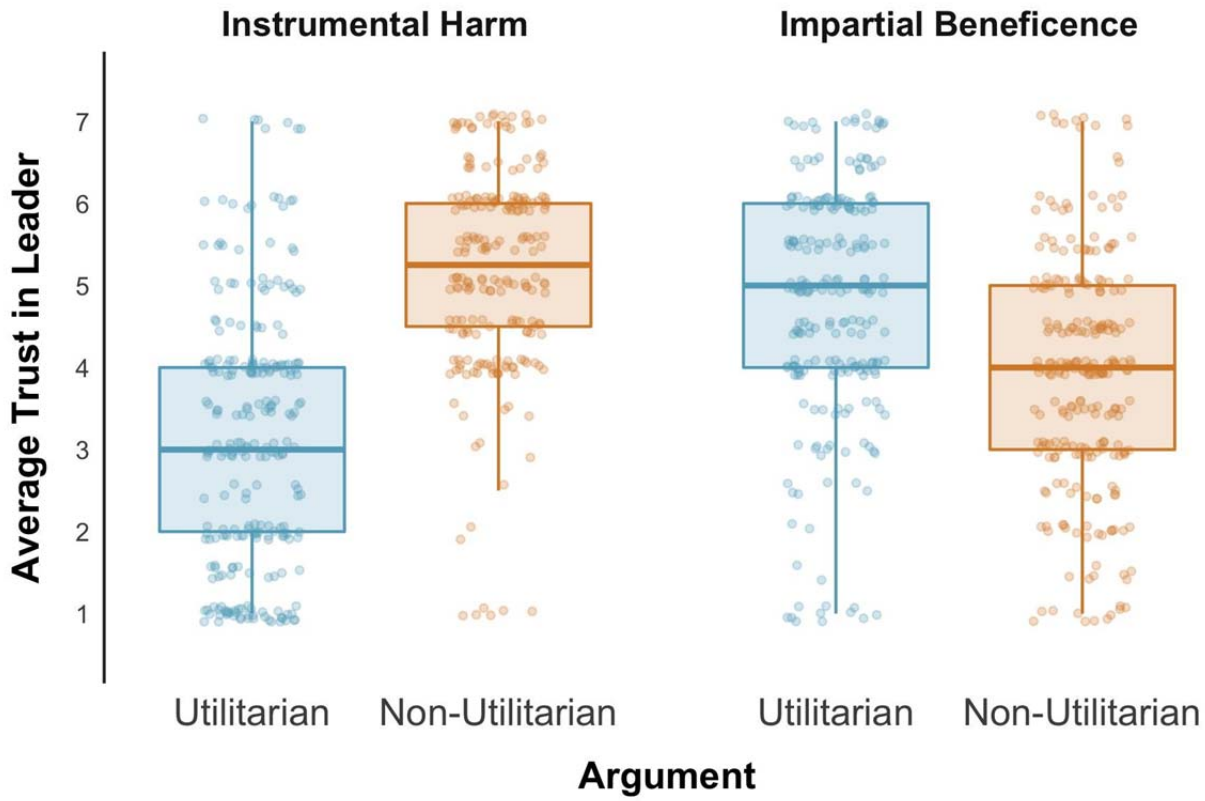
Supplementary Figure 1. Impact of the COVID-19 Pandemic on Countries of Intended Recruitment. (A) Number of confirmed COVID-19 cases per 100 thousand people in each country of intended recruitment. (B) Number of COVID-19 deaths per 100 thousand people in each country of intended recruitment. (C) Absolute number of confirmed COVID-19 cases in each country of intended recruitment. (D) Absolute number of COVID-19 deaths in each country of intended recruitment. COVID-19 confirmed cases and death rates were taken from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University[74] (last update: October 18th, 2020). Population estimates for each country were taken from the United Nations' World Population Prospects (last update: July 1st, 2019).



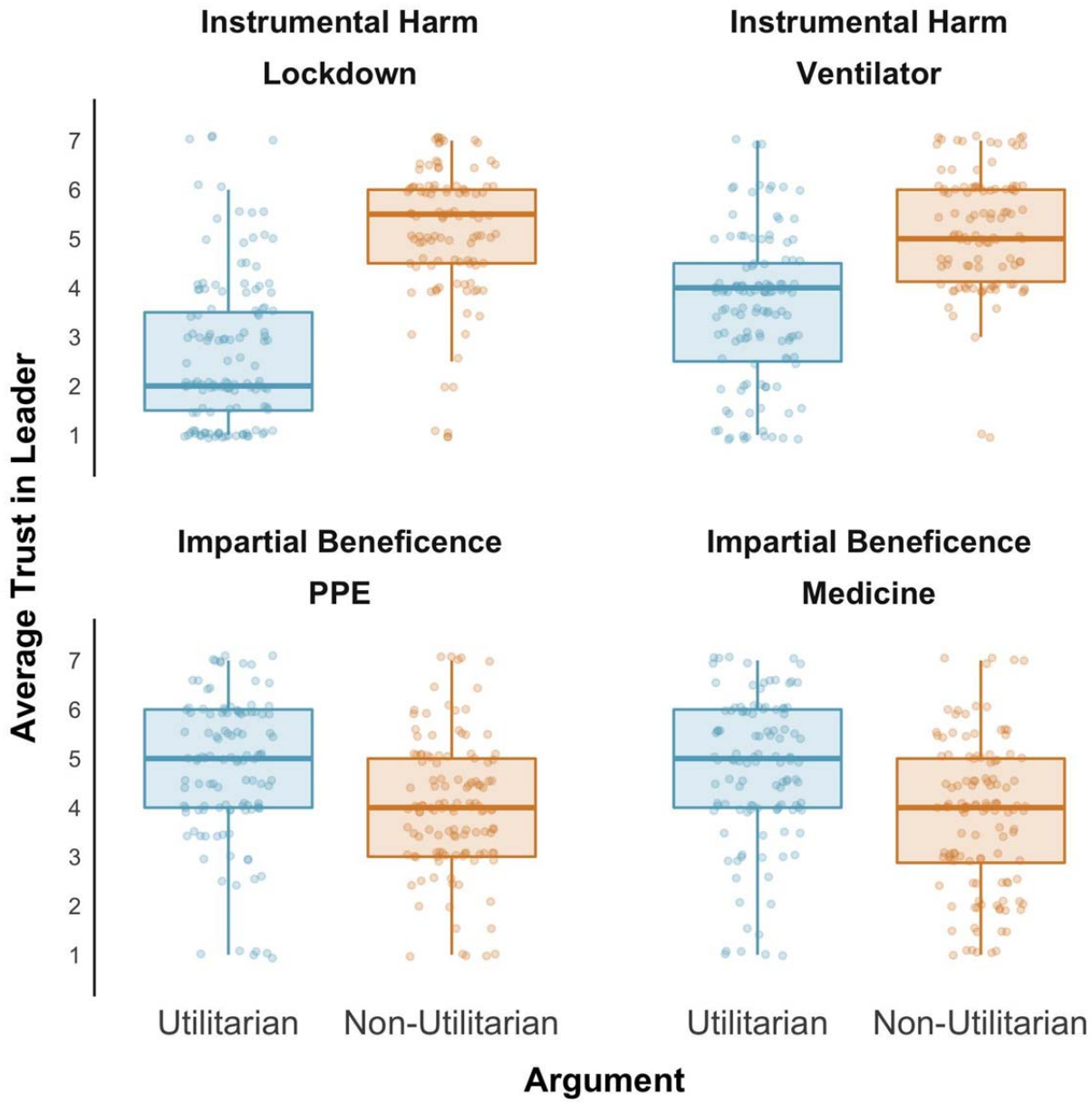
Supplementary Figure 2. ***Self Reported Trust in Pilot 1.*** Average self-reported trust in utilitarian vs non-utilitarian leaders in Pilot 1 (N = 98), separately for Instrumental Harm dilemmas (Lockdown, Tracing, and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders were seen as more trustworthy than utilitarian leaders for Instrumental Harm dilemmas (B = -1.18, SE = 0.10, t(875) = -11.72, p < .001, CI = [-1.37, -0.98]), while the reverse was observed for Impartial Beneficence dilemmas (B = 1.15, SE = 0.12, t(875) = 9.37, p < .001, CI = [0.91, 1.39]). Bars correspond to median scores, lower and upper hinges correspond to the first and third quartiles, respectively, and whiskers ends correspond to the most extreme data points within 1.5 times the interquartile range.



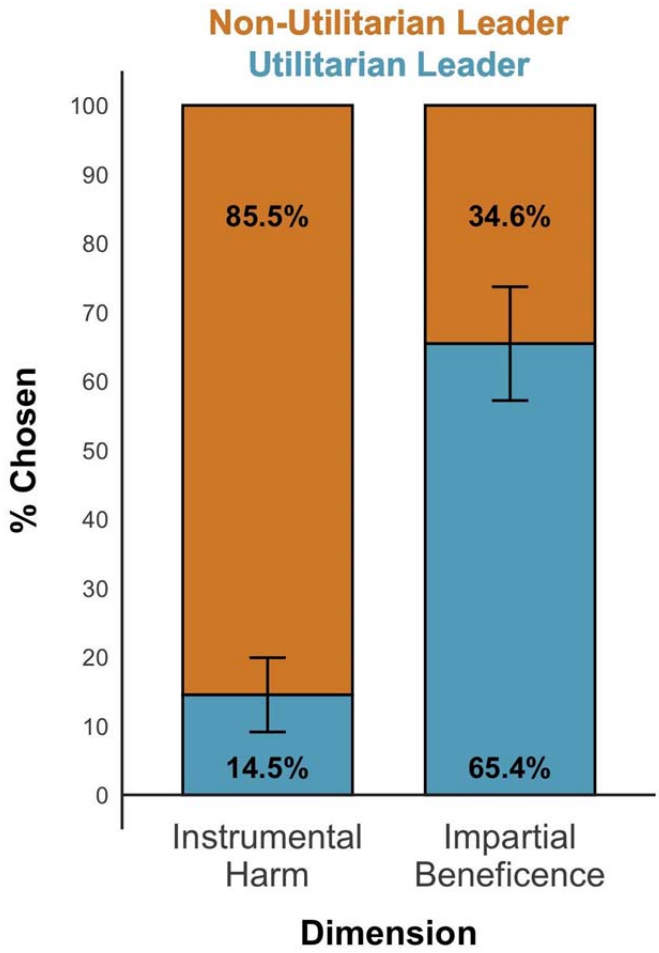
Supplementary Figure 3. ***Self Reported Trust by Dilemma in Pilot 1.*** Average self-reported trust in utilitarian vs. non-utilitarian leaders in Pilot 1, separately for each dilemma, including both Instrumental Harm dilemmas (Lockdown, Ventilators, and Tracing) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders were seen as more trustworthy than utilitarian leaders in both Instrumental Harm dilemmas, but not in either Impartial Beneficence dilemmas. Bars correspond to median scores, lower and upper hinges correspond to the first and third quartiles, respectively, and whiskers ends correspond to the most extreme data points within 1.5 times the interquartile range.



Supplementary Figure 4. ***Self Reported Trust in Pilot 2***. Average self-reported trust for the utilitarian vs. non-utilitarian leaders in Pilot 2 (N = 469), separately for Instrumental Harm dilemmas (Lockdown and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders were seen as more trustworthy than utilitarian leaders in Instrumental Harm dilemmas (B = -2.02, SE = 0.17, $t(454)$ = -11.59, $p < .001$, CI = [-2.36, -1.68]), but not in Impartial Beneficence dilemmas (B = 0.86, SE = 0.17, $t(455)$ = 5.00, $p < .001$, CI = [0.52, 1.19]). Bars correspond to median scores, lower and upper hinges correspond to the first and third quartiles, respectively, and whiskers ends correspond to the most extreme data points within 1.5 times the interquartile range.



Supplementary Figure 5. Self Reported Trust by Dilemma in Pilot 2. Average self-reported trust in utilitarian vs non-utilitarian leaders in Pilot 2, separately for each dilemma, including both Instrumental Harm dilemmas (Lockdown and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders were seen as more trustworthy than utilitarian leaders in both Instrumental Harm dilemmas, but not in either Impartial Beneficence dilemmas. Bars correspond to median scores, lower and upper hinges correspond to the first and third quartiles, respectively, and whiskers ends correspond to the most extreme data points within 1.5 times the interquartile range.



Supplementary Figure 6. ***Voting Choices in Pilot 2***. Model estimates of the percentage of participants who chose to entrust utilitarian vs. non-utilitarian leaders in the voting task in Pilot 2 (N = 452), separately for Instrumental Harm (Lockdown and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders were more likely to be voted in Instrumental Harm dilemmas, but not in Impartial Beneficence dilemmas ($B = 2.41$, SE = 0.33, $z = 7.30$, $p < .001$, CI = [1.77, 3.13], OR = 11.13). Error bars represent standard error of the model estimates.