

Moral reasoning about the COVID-19 crisis

Authors: Joaquin Navajas^{1,2}, Facundo Álvarez Heduan¹, Gerry Garbulsky¹, Enzo Tagliazucchi^{2,3}, Dan Ariely⁴, & Mariano Sigman^{1,2,5}

¹ School of Business, Universidad Torcuato Di Tella, Av. Figueroa Alcorta 7350, Buenos Aires, C1428BCW, Argentina

² National Scientific and Technical Research Council (CONICET), Godoy Cruz 2290, Buenos Aires, C1425FQB, Argentina

³ Physics Department, Universidad de Buenos Aires, Av. Intendente Guiraldes 2160, Buenos Aires, C1428EGA, Argentina.

⁴ The Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, USA

⁵ Facultad de Lenguas y Educación, Universidad Nebrija, Calle de Sta. Cruz de Marcenado 27, Madrid 28015, Spain

Corresponding author: *Joaquin Navajas*, e-mail: joaquin.navajas@utdt.edu

Abstract

The COVID-19 pandemic has raised complex moral dilemmas that have been the subject of extensive public debate. In this paper, we study the reasoning mechanisms used by individuals when they judge the acceptability of a set of controversial actions: relaxing data privacy standards to allow public control of the pandemic, forbidding public gatherings until the pandemic is over, denouncing a friend who violated COVID-19 protocols, prioritizing younger over older patients when medical resources are scarce, and reducing animal rights to accelerate vaccine development. We collected acceptability judgements in 10 Latin American countries (N=15,420) and performed a pre-registered replication in the United States (N=1,300). A formal analysis of the intrinsic correlations between responses to different dilemmas suggests that people consistently used two reasoning processes: one that maximizes human life expectancy and one that cares about the health of all sentient lives in an equitable manner. The extent to which people employ these considerations varies across individuals but can be robustly predicted by a standardized moral reasoning scale. Moreover, we show that people's reliance on these two arguments correlates with the severity of the pandemic in their country. Altogether, these findings elucidate the reasoning processes underlying moral reactions to the COVID-19 crisis.

Introduction

Over the last years, the psychological foundations of moral decision-making have been extensively studied by looking at how people react to abstract dilemmas (1-7). For example, in trolley-type scenarios, participants are asked to decide whether it is acceptable to perform an action that will save the lives of some people at the cost of killing others (1). While remarkably informative, these theoretical scenarios have two main limitations. First, people may express preference for the same decision while relying on fundamentally different reasoning mechanisms (8). Second, people are imprecise at estimating their future emotional states (9-12) and, as a consequence, they may be inaccurate at forecasting how they will judge and react when moral dilemmas are presented in the heat of the moment (13-16).

The COVID-19 pandemic has posed the world with an extraordinarily difficult challenge that has brought moral dilemmas to the public sphere. Issues such as how to assign scarce medical resources (e.g., ventilators in cases of limited availability (17, 18)), or whether it is acceptable to share sensitive private data to effectively trace the virus (19), or the tension between ensuring physical distancing or allowing economic and social activities (20), have shifted moral reasoning from theoretical to practical considerations and became part of public deliberations all the way from lay people to policy makers.

This results in an unprecedented opportunity to understand how people reason about moral decisions, which are relevant to policy making, in the midst of a healthcare crisis. Previous research has shown that moral dilemmas can help providing a graded notion of values across peoples and cultures (21). But, above and beyond quantifying average preferences, here we aimed at performing a formal analysis of the intrinsic correlations between moral judgements across different scenarios, which in turn can serve to infer the reasoning mechanisms that result in moral judgements. Summarizing, the main objective of this work is to unfold the processes of thought underlying moral preferences during the emergence of real-life moral discussions.

We ground our work on the classic partition of moral reasoning between utilitarian and deontological principles (22). Utilitarianism is a normative theory that advocates that decisions ought to maximize a utility function reflecting the happiness and well-being of all affected individuals. As an example, in trolley-like dilemmas where different decisions lead to outcomes that vary in the number of deaths, the utilitarian prescription is to select the option in which less people die. Deontology is a different normative theory that proposes to judge the morality of a decision based on the action itself rather than on their consequences. The standard deontological decision in trolley-type scenarios is to do nothing so to avoid being in the position of deciding who dies and who does not.

More recently, it has been argued that this one-dimensional organization of moral thinking, with utilitarianism in one extreme and deontologism in the other, is over-simplistic and that those engaging in utilitarian thinking based on trolley-type dilemmas may in fact be relying on different processes of thought that produce the same outcome (8). According to this view, utilitarian preferences result from two very distinct aspects: a ‘negative’ dimension that reflects a permissive attitude towards *instrumental harm*, and a ‘positive’ dimension, called *impartial beneficence*, which reflects an unbiased concern for the wellbeing of all sentient lives. Previous research has shown that, while *instrumental harm* correlates with psychopathic tendencies, *impartial beneficence* is associated with higher empathic concern (8).

We put different theories of utilitarian thinking to test and inquire which models’ predictions are consistent with the preferences expressed by people considering moral decisions about the pandemic. To address this aim, we designed five scenarios that probe different ways of moral reasoning in contemporary dilemmas (**Figure 1A**). The first three dilemmas pose a tension between public health and other values of wellbeing: 1) effectively tracing the virus to control its spread versus protecting sensitive personal data, 2) ensuring physical distancing by forbidding public gatherings and business operations versus allowing those activities before a vaccine is developed, and 3) notifying a COVID-19 protocol breach versus protecting a close friend from facing prison. The remaining two dilemmas ask whether

one should give priority to some lives above others in public health. The fourth dilemma is about the assignment of ventilators in case of limited availability and asks whether all patients should be treated equally or if younger individuals should be prioritized. This scenario could be thought as a real-life variant of a trolley-dilemma where one has to choose between saving younger or older people (21). The fifth dilemma poses a tension between respecting the rights of animals in medical research versus accelerating the development of a vaccine.

One way to characterize moral responses to these scenarios is by looking at majority or average opinions. However, in order to understand the reasoning mechanisms underlying this highly heterogeneous set of dilemmas, one needs to estimate the pattern of correlations between different judgements. For example, if a person thinks that younger patients should be prioritized in their access to ventilators, will this person be more likely to support restrictions on social and economic activity? To answer this question, and to identify the different prototypical ways of thinking triggered by these scenarios, here we examine the dimensionality of moral responses (**Figures 1B** and **1C**).

This analysis requires a substantial amount of data, much more than what is typically obtained in classic psychological and cognitive experiments. To solve this, we capitalized on a program which we have developed over the last years to perform cognitive experiments with large audiences (23-27). Through this program we could generate a sample of more than 15,000 participants proceeding from 10 Spanish-speaking Latin American countries (Study 1, $N = 15,420$, 60.1% female, mean age: 32.0 y.o., range 18-95 years). We also tested the robustness of our findings in a second pre-registered study with a representative sample of the United States (Study 2, $N = 1,300$, 48.3% female, mean age: 41.2 y.o., range 18-89 years).

Participants read the five scenarios about the COVID-19 pandemic in randomized order as well as the classic trolley problem in either its personal or impersonal version (see **Methods** for details). Each scenario described a clear action and participants reported three quantities using sliders: their perceived acceptability of that action (in a scale that ranged from

‘completely wrong’ to ‘completely right’), their confidence in that judgement (from ‘completely unsure’ to ‘completely sure’) and the amount of distress that they think they would feel if they had to make a decision in that context (from ‘none’ to ‘a great deal’).

To distinguish between unidimensional (28) and bidimensional (8) theories of moral reasoning, one needs to infer the underlying dimensionality of the distribution of acceptability judgements. Responses to these dilemmas, which varied widely across the population (**Figure S1**), define a native 5-dimensional space but, intuitively, if all points are scattered across a line, the underlying dimensionality of the data is one, if they are scattered within a plane, the dimensionality is two, and so on so forth (**Figure 1B**). While there are various techniques to formalize this intuition, one simple approach to statistically test the underlying dimensionality of the data is to perform a principal component analysis (PCA).

The advantage of this analysis, beyond its simplicity, is that it also provides the principal directions across which people tend to be aligned, which may allow us to identify -to a first degree of approximation- the argument of utility functions. To infer the elements of wellbeing that are maximized by people who endorse utilitarian principles, we study the correlation between individual differences in utilitarian tendencies and the participants’ projection onto the principal components. Here, we consider and present two hypotheses about the relationship between utilitarian thinking and moral responses to the pandemic.

The first hypothesis is that utilitarian thinking is associated with a focus on maximizing *human life expectancy*. If that were the case, then people prioritizing public health in the first three dilemmas should also prioritize human over non-human animal life and younger over older patients (left panel of **Figure 1C**). The second hypothesis is that utilitarian thinking should correlate with a focus on *equitable public health*, i.e., people endorsing utilitarian principles should prioritize public health in a fair and impartial manner contemplating and respecting all lives. The prototypical response predicted by this model is prioritizing public health in the first three dilemmas while respecting the rights of older patients and non-human

animals in the last two scenarios (left panel of **Figure 1C**). While these two hypotheses are mutually exclusive according to unidimensional models of utilitarian reasoning, they could in principle co-exist in bidimensional models.

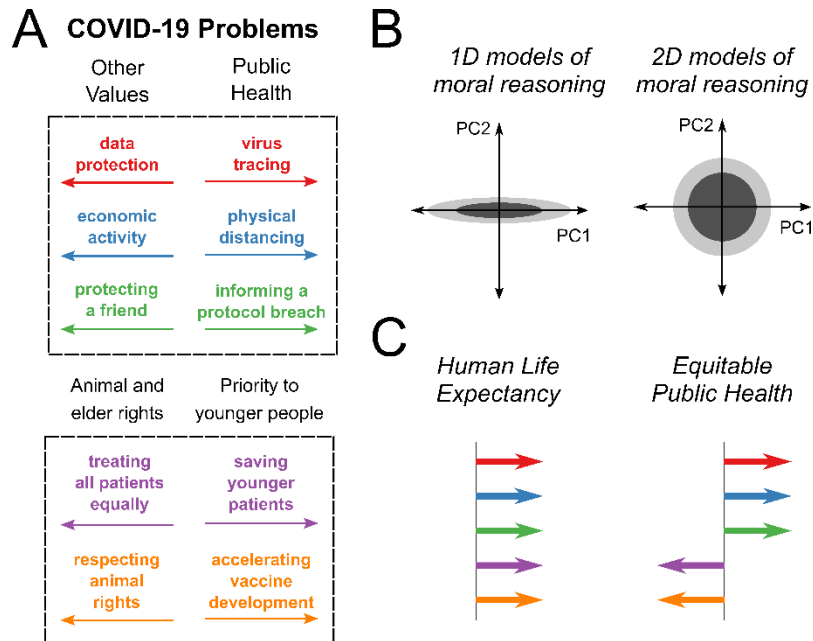


Figure 1. Moral problems about the COVID-19 pandemic and theoretical predictions. **A)** We designed five moral scenarios that relate to problems of the COVID-19 outbreak (see **Methods** for full descriptions of each scenario). The first three pose a tension between prioritizing public health or other values and aspects of wellbeing. The last two ask if it is acceptable to prioritize younger human lives at the cost of breaking elder and animal rights. **B)** Unidimensional models of moral thinking predict that responses to these dilemmas should project onto a single principal component reflecting the reliance on deontological or utilitarian reasoning. Bidimensional theories, instead, suggest that moral responses should be better explained by two components, each one reflecting a different type of utilitarian reasoning processes. **C)** Dilemmas were conceived to examine the hypothesis of two distinct focuses on public health across. The first is a concern about maximizing human life expectancy whereby individuals prioritizing public health over other aspects of wellbeing also prioritize younger human life over older and animal lives (left panel). The second hypothesis is a focus on equitable public health where prioritization to public health is compatible with respecting elder and animal rights (right panel).

Results

Individual differences in utilitarianism

We measured individual differences in utilitarian thinking (8) through a standardized scale (i.e., the Oxford Utilitarianism Scale, OUS, see **Methods** for details). This scale allows measuring individual variations in overall utilitarianism as well as in each of the two dimensions of utilitarian thinking (OUS-IH: instrumental harm and OUS-IB: impartial beneficence). The maximum degree of overall utilitarianism is achieved when people score highly on both sub-scales (**Figure 2A**, upper-right quadrant, *utilitarians*). On the opposite end of the spectrum, individuals scoring lowly on both sub-scales are identified as *deontologists* (**Figure 2A**, lower-left quadrant). Using a mean split on both sub-scales we found that 26.6% ($n = 4,099$) of the individuals in our sample were classified as utilitarians and 27.5% ($n = 4,242$) of them as deontologist. We also observed a large fraction of participants who scored highly on instrumental harm but lowly on impartial beneficence (i.e., *instrumentalists*, upper-left quadrant, 22.8%, $n = 3,517$) and vice versa (i.e., *impartialists*, lower-right quadrant, 23.1%, $n = 3,562$).

The study presenting this scale found a low degree of collinearity between both dimensions of utilitarianism (i.e., $r=0.14$, $p<.01$, see Table 6 in ref. (8)). Here, we also observed a low correlation between these two variables ($r=0.13$, $p<10^{-64}$) justifying the use of both dimensions of utilitarianism as separate regressors ($R^2=0.01$, $VIF=1.01$). We also replicated a key finding of the original study: individuals with high scores on any of the two dimensions were more likely to support utilitarian decisions in the trolley dilemma (i.e., agree to sacrifice the life of a person to save a greater number of individuals) both in the impersonal (**Figure 2B**, correlation with Instrumental Harm: $r=0.26$, $p<10^{-128}$, correlation with Impartial Beneficence: $r=0.11$, $p<10^{-21}$) as well as in the personal (**Figure 2C**, correlation with Instrumental Harm: $r=0.39$, $p<10^{-281}$, correlation with Impartial Beneficence: $r=0.10$, $p<10^{-19}$) variants. Also consistent with that same previous study, we found that the correlation of

utilitarian judgements in trolley-type dilemmas with instrumental harm was higher than the correlation with impartial beneficence (Test for equal correlations, $z=10.39$, $p<10^{-200}$). Altogether, these results confirm that classic trolley dilemmas intermingle these two factors that contribute to utilitarian reasoning.

The bidimensional hypothesis of utilitarian thinking makes two important predictions on other variables of the trolley-problem which we tested in this large-scale sample. The first one relates to confidence: if impartial beneficence and instrumental harm are separate reasoning processes, then people's confidence should vary with the extremity of both two sub-scales. We observed strong evidence consistent with this prediction (bi-variate regression of absolute z-scored OUS sub-scales on confidence in the trolley problem, effect of instrumental harm: $\beta = .22\pm.02$, $p < 10^{-28}$, effect of impartial beneficence: $\beta = .13\pm.02$, $p < 10^{-11}$, see **Figure S2**). This demonstrates that people scoring highly or lowly on instrumental harm or impartial beneficence displayed higher confidence in their judgements.

The second prediction relates to the self-reported level of distress about making these decisions. According to previous research, sub-clinical psychopathy has been shown to correlate positively with instrumental harm and, instead, empathic concern correlates positively with impartial beneficence but negatively with instrumental harm (8). This predicts that self-reported distress will be greater for those people with more impartial beneficence and lesser for those with more instrumental harm. Observed values of self-reported distress about the trolley problem provided strong evidence supporting the bidimensional nature of both sub-scales (bi-variate regression of z-scored sub-scales on distress about the trolley problem, effect of instrumental harm: $\beta = -.15\pm.01$, $p<10^{-41}$, effect of impartial beneficence: $\beta = .14\pm.01$, $p<10^{-35}$, see **Figure S2**).

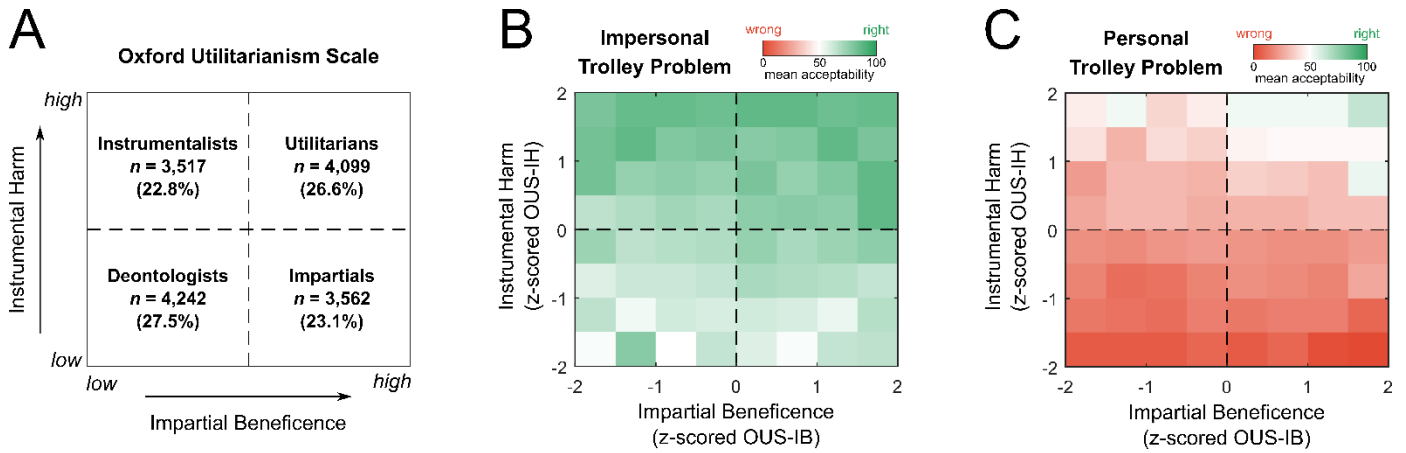


Figure 2. Instrumental harm and impartial beneficence as different dimension of utilitarian reasoning. A) The Oxford Utilitarianism Scale (OUS) provides an overall metric of utilitarianism as well as two sub-scales, each one measuring the reliance on different aspects of utilitarian thinking: impartial beneficence (OUS-IB, x-axis) and instrumental harm (OUS-IH, y-axis). According to this model, *deontologists* are those who score lowly on both sub-scales (lower-left quadrant), *impartials* score highly on OUS-IB but not on OUS-IH (lower-right quadrant), *instrumentalists* score highly on OUS-IH but lowly on OUS-IB (upper-left quadrant), and those scoring highly on both scales are classified as *utilitarians* (upper-right quadrant). Each quadrant describes the number and proportion of participants in our sample. **B-C)** Images display the tendency to accept the utilitarian decision in the impersonal (B) and personal (C) versions of the trolley problem, as a function of instrumental harm and impartial beneficence. Colour codes the mean acceptability rating averaged across participants at a given level of z-scored OUS-IB (x-axis) and OUS-IH (y-axis).

The overarching goal of this study is to understand the structure of moral reasoning in real-life dilemmas that have emerged during the COVID-19 crisis. We organize and narrow this general and broad aim by asking three questions that examine different fundamental fingerprints of the data: 1) What is the underlying dimensionality of the space of moral responses?; 2) What are the relevant principal components across which the data is organized?; and 3) How do these dimensions relate to individual differences in utilitarian thinking? To answer these questions, we analyzed the acceptability ratings that participants provided in response to the five moral scenarios about the COVID-19 pandemic (**Figure 3A**).

The first question can be naturally addressed by investigating the number of significant principal components (i.e., the number of dimensions) that account for variance in the data above and beyond what would be expected by pure chance (see **Methods**). This analysis unveiled two main components explaining a significant proportion of the variance (**Figure 3B**, random permutation test with 10,000 simulations, $p < 10^{-5}$), which is a mathematical way of reflecting that the embedding of moral responses is two-dimensional.

To address the second question, we looked at the directions of these two principal components, i.e. which combinations of responses across dilemmas explain for large fractions of the variance in the data (**Figure S3**, information displayed in a PCA biplot (29) along with the joint probability distribution of the two components). This analysis revealed that the first principal component corresponds to a concern about *human life expectancy* (upper-left panel of **Figure 3B**), namely, a pattern of stereotyped responses where people prioritizing public health in the first three dilemmas also prioritize younger over older patients and human over non-human life. The second component is consistent with a focus on *equitable public health* (upper-right panel of **Figure 3B**) whereby prioritizing public health in the first three dilemmas is associated with the tendency to respect ethical procedures of animal research and the rights of elderly patients.

The third question asks whether these patterns of correlations match with the two constituting elements of utilitarianism (8). Do these dimensions correspond to people relying on their tolerance to instrumental harm or their concern for impartial beneficence? To address this, we performed a regression of both components with the two sub-scales as predictor variables (**Figures 3C and 3D**). We found that the first principal component (**Figure 3C and Figure S4**) was significantly modulated by instrumental harm ($\beta = .31 \pm .04$, $p < 10^{-12}$) but not by impartial beneficence ($\beta = .02 \pm .03$, $p = .64$). Conversely, the second principal component (**Figure 3D and Figure S4**) was explained by individual variations in impartial beneficence ($\beta = .24 \pm .03$, $p < 10^{-12}$) but not in instrumental harm ($\beta = -.05 \pm .04$, $p = .23$). These two findings are consistent with the prediction that different types of utilitarian thinking differ in the way they react to dilemmas which set priorities to some or all lives. Individuals with high scores of impartial beneficence will generally prefer not to assign priorities on ventilator use and will not agree on lowering the threshold of animal rights to accelerating the development of a vaccine, while the opposite choices will be preferred by those scoring highly on instrumental harm. Our results also reveal that, while these two different utilitarian styles seem to be maximizing different utility functions, they both agree on prioritizing public health over other non-health related aspects of wellbeing such as data privacy, economic activity, and friendship (i.e. the first three scenarios in **Figure 1A**).

Pre-registered replication study in the United States

One potential concern with Study 1 is that we could not control the representativeness of the sample that we obtained. For this reason, we performed a second pre-registered study in a different setting (Study 2), using a sample that is representative of the United States population in terms of age, gender, and ethnicity (**Table S1**). To determine sample size, we performed a Monte Carlo power analysis based on the data collected in Study 1 (**Figure S5**). Based on this analysis, we recruited 1,300 participants, yielding an estimated power of 83.8%.

A second fundamental motivation for conducting this study is that it allowed us testing the robustness and replicability of our findings. With this aim, we pre-registered the hypotheses and methods of this study, specifying that its aim was to replicate the three main findings observed in Study 1 (<https://aspredicted.org/59kk7.pdf>). First, we predicted the existence of two principal components explaining a significant proportion of the variance in the data (**Figure 3B**). Second, we hypothesized that these two dimensions would reflect a concern for human life expectancy (**Figure 1C**, i.e. a principal component with positive loadings for all five scenarios) and an interest in equitable public health (**Figure 1C**, i.e. a principal component with positive loadings for the first three scenarios and negative loadings for the last two). Third, we predicted a positive correlation between instrumental harm and human life expectancy and a positive correlation between impartial beneficence and equitable public health (**Figures 3C and 3D**).

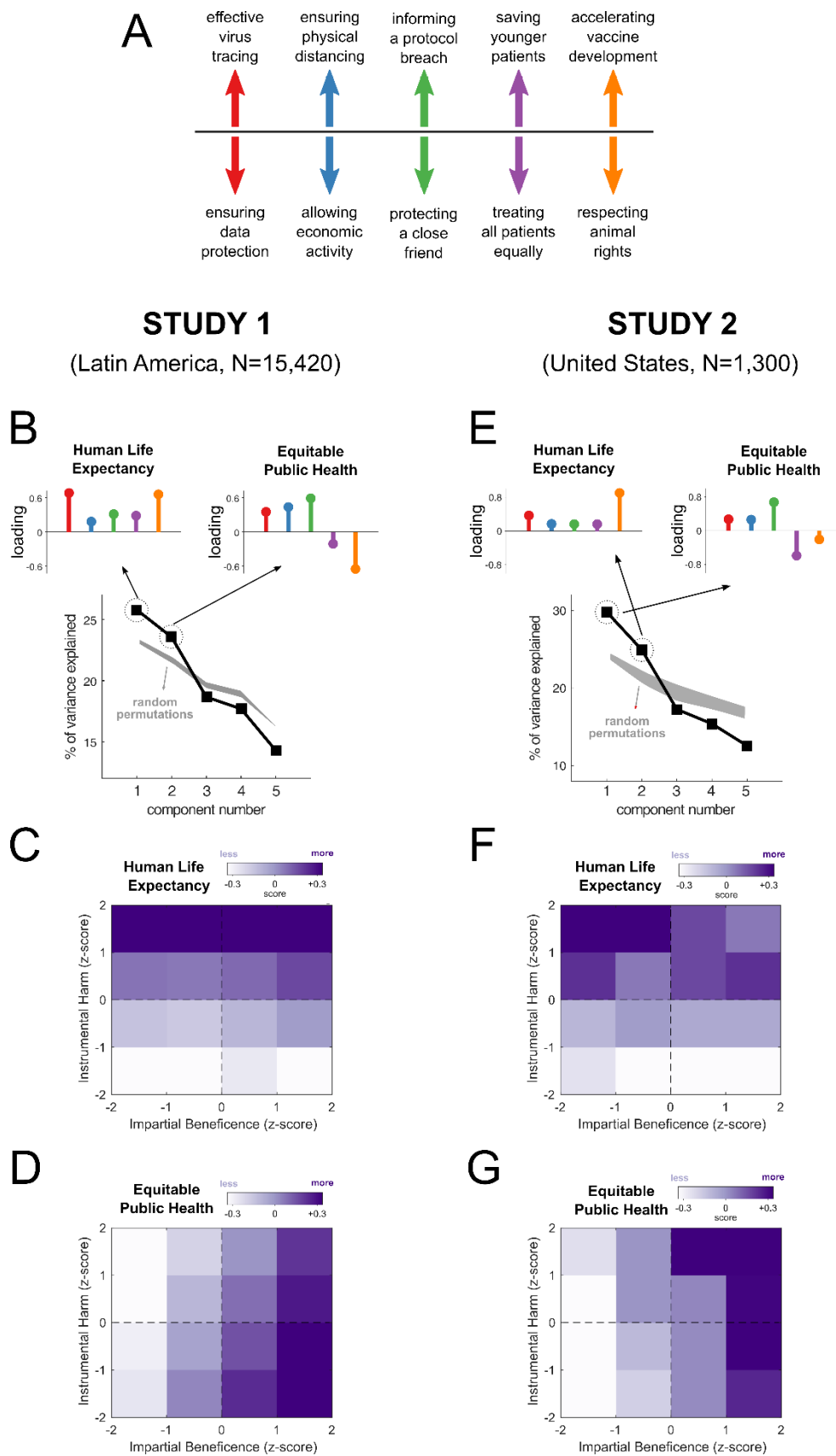
We first focused on the prediction that only two dimensions would explain a significant proportion of the variance in the data. To formally test this, we performed a random permutation analysis. As per our pre-registration, we generated 10,000 surrogated datasets by randomly shuffling the participants' labels for each scenario. Applying PCA on each surrogate, we estimated the distribution of explained variances for each component under the null hypothesis that there are no correlations in the data. As in Study 1, we observed that only the first two principal components explained more variance in the data than what is expected by chance (**Figure 3E**, random permutation test, $p < 10^{-5}$).

Our second prediction was that these two dimensions would reflect human life expectancy and equitable public health. We tested this hypothesis by examining the signs of the five loadings for the first two principal components. As predicted, we found that one of those two dimensions had positive loadings for all dilemmas (upper-left panel of **Figure 3E**, a dimension consistent with a concern for human life expectancy) and the other component had positive loadings for the first three dilemmas and negative loadings for the last two (upper-right panel of **Figure 3E**, a dimension which reflects an interest in equitable public health).

The only difference between these findings and the ones observed in Study 1 is that the order of the two principal components was reversed (i.e., equitable public health explained more variance than human life expectancy). However, it is important to remark that we explicitly did not make any pre-registered claim about which of the two dimensions would be more representative in the data. This is because we reasoned that the different contexts associated to both studies may switch on how people give priority to these dimensions. Instead, the prediction of our study was that, while this may vary, variance in the data would still be structured in these two underlying dimensions. This effect, which was first observed in Study 1 (**Figure 3B**), was successfully replicated in Study 2 (**Figure 3E**).

Our third and last prediction was that each of these two components should correlate with a different aspect of utilitarian thinking, as quantified by the two sub-scales of the Oxford Utilitarianism Scale. As predicted, we observed that the component reflecting a concern for human life expectancy was positively correlated with having a permissive attitude towards instrumental harm (**Figure 3F**, $r=.20$, $p<10^{-13}$). We also found a significant correlation between the component suggesting an interest in equitable public health and impartial beneficence (**Figure 3G**, $r=.39$, $p<10^{-49}$). Importantly, these correlations were positive and significantly higher than the opposite and corresponding associations, i.e., if we correlated human life expectancy with impartial beneficence (Test for Equal Correlations, $z=5.5$, $p<10^{-6}$) or equitable public health with instrumental harm ($z = 4.8$, $p<10^{-6}$).

Overall, these findings indicate that Study 2 successfully replicated the results obtained in Study 1 despite using a different sampling strategy and collecting data in a different language, country, and moment of the pandemic. These results provide strong evidence that moral judgements in response to these five dilemmas are organized in two stereotypical ways of reasoning, which are well predicted by people's scores in impartial beneficence and instrumental harm.



(Figure legend in next page)

Figure 3. Dimensionality analysis of moral problems of the pandemic and relationship with utilitarian reasoning processes. **A)** We performed a principal component analysis (PCA) of the acceptability ratings provided by participants in response to five moral dilemmas about the COVID-19 crisis. The arrows indicate how we coded the loadings of the principal components. For example, a positive loading in red reflects more agreement with public officials using private data to effectively trace the virus. **B)** In Study 1 (performed in Latin America with 15,420 participants), we observed that the first two principal components explained a significant amount of variance in the data. Black squares show the percentage of variance explained by each of the five principal components, and insets display the loadings of each variable, with each color coding a different dilemma as in panel A. The first principal component is consistent with a concern about human life expectancy and the second component with a focus on equitable public health. The light grey shade shows the ranges of variance percentages explained by 10,000 random permutations of the data. **C)** Relationship between different dimensions of utilitarian reasoning and the first principal component (i.e., human life expectancy). Images show the average score of each component for participants with a given value of z-scored impartial beneficence (OUS-IB, x-axis) and instrumental harm (OUS-IH, y-axis). The data suggests that the principal component reflecting human life expectancy show greater correlation with instrumental harm than with impartial beneficence. **D)** Same as C but for the second principal component. The data suggests that the principal component reflecting equitable public health show greater correlation with impartial beneficence than with instrumental harm. E-F) Same as B-D but for the pre-registered replication study performed in the United States (Study 2).

Contextual and personality factors

Our next question was whether, above and beyond these effects, any residual variance in the data could be explained by contextual (10, 13) and personality (30, 31) factors that may shape real-life moral decisions at constant values of utilitarian scores. This effort can be seen as a way of putting together different aspects of an individual -their utilitarian fingerprints, their personality traits, and contextual elements- to disentangle how they differentially affect moral responses to the pandemic. To this end, we focused on the large-scale dataset obtained in Study 1 and ran a multivariate regression on the projection of each principal component as a function of 15 moral, contextual, personality, and demographic variables (see **Methods**, **Table S2** and **Table S3**).

Previous research has suggested that negative contexts trigger affective states that could modify moral preferences (13), and so we reasoned that, even at constant values of utilitarianism, individuals proceeding from countries where the COVID-19 crisis is currently more severe could display greater preference to prioritize public health in moral dilemmas. To test this hypothesis, we analyzed the data collected in Study 1 with participants proceeding from 10 Latin American countries (**Figures 4A** and **4B**). Importantly, by the time when the study was performed (May 2020), the impact of the pandemic had shown a thirty-fold difference in confirmed COVID-19 deaths per million people (i.e., 6 for Uruguay and over 180 for Ecuador).

We observed that the intensity of the crisis, quantified by the per capita number of COVID-19 deaths, significantly modulated the participants' projection onto the first ($\beta = .05 \pm .01$, $p < 10^{-4}$) and second ($\beta = .07 \pm .01$, $p < 10^{-8}$) principal components (**Figure 4C**). This association remained significant if, instead of using the per capita number of deaths, we used the per capita number of confirmed COVID-19 cases ($\beta = .03 \pm .01$, $p = .002$ for the first principal component, $\beta = .04 \pm .01$, $p = .001$ for the second principal component). Overall, this analysis

suggests that variations in the severity of the context modulated the projection onto both principal components.

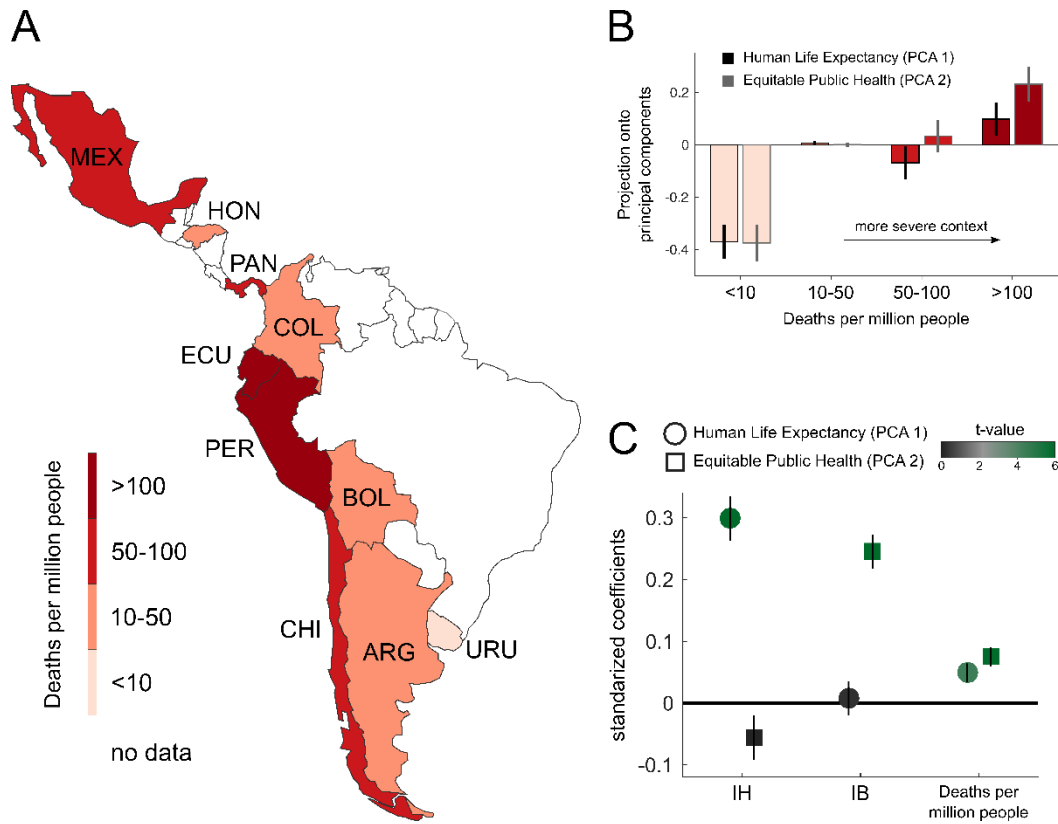


Figure 4. Country-level effects on moral preferences. **A)** In Study 1, we collected data from 10 countries where the impact of the pandemic has been highly dissimilar at the time when the study was performed (May 2020). The map displays the countries from where we have obtained data. Colors show the per-capita total number of confirmed COVID-19 deaths by the end of data collection. Country labels are Uruguay (URU), Argentina (ARG), Colombia (COL), Honduras (HOR), Bolivia (BOL), Chile (CHI), Mexico (MEX), Panama (PAN), Peru (PER), and Ecuador (ECU). **B)** Projection onto the first two principal components (PCs) for participants proceeding from different countries. Participants were grouped based on the number of COVID-19 deaths per million people at the country from where they proceed. Less than 10: URU. Between 10 and 50: ARG, BOL, COL, and HON. Between 50 and 100: CHI and MEX. More than 100: PER and ECU. Bars denote average PC score across participants, and vertical lines depict SEM. **C)** We performed a multivariate regression analysis on the scores obtained for each PC. Circles show coefficient estimates for the first principal component (Human Life Expectancy) and squares show coefficient estimates for the second principal component (Equitable Public Health). Predictor variables displayed are Instrumental Harm (IH), Impartial Beneficence (IB) and the country-level pandemic's severity, quantified as the COVID-19 deaths per million people. Vertical lines show SEM and colors code the t-value associated to each coefficient. For the full list of predictor variables included in the analysis, see Tables S2 and S3.

However, the correlation of COVID-19 deaths (or any other variable, of course) with a principal component does not imply directly a correlation with responses to individual dilemmas. Just to give a simple example: if 4 of the 5 dilemmas would correlate positively with COVID-19 deaths then a principal component with all positive loadings (like the first principal component in Study 1) may also show a significant correlation although one of the dilemmas that defined this component did not correlate with COVID-19 deaths. Therefore, to understand the relationship between the severity of the pandemic and people's responses to individual dilemmas, we performed a post-hoc correlation analysis. We found that the correlation between the number of COVID-19 deaths and the acceptability ratings to the first three scenarios was significantly positive (Scenario 1: $r=.03$, $p=.001$; Scenario 2: $r=.03$, $p=.001$; Scenario 3: $r=.04$, $p=7\times 10^{-7}$) while we observed a non-significant correlation with the responses to the last two scenarios (Scenario 4: $r=.01$, $p=.14$; Scenario 5: $r=.003$, $p=.63$). This finding suggests that the observed modulation given by the severity of the pandemic on the two principal components is driven by the first three scenarios, which have positive loadings for both components. In other words, the association between the severity of the pandemic and the two principal components is a result of a stronger prioritization of public health over other aspects of wellbeing and is unrelated to the trade-off between younger and older patients (Scenario 4) or between human and animal rights (Scenario 5).

To further examine the potential role played by contextual factors, we asked participants to report their personal proximity to the pandemic by indicating whether they were diagnosed with COVID-19 and/or if they knew someone who did. In our sample, 0.6% ($n=92$) reported to have tested positive for COVID-19, 31.5% ($n=4,949$ participants) reported to have at least one acquaintance who tested positive, and 67.9% ($n=10,685$ participants) reported not knowing anyone diagnosed with the virus. We found no evidence that being diagnosed with COVID-19 (first principal component: $\beta = -.06\pm.05$, $p=.24$, second principal component: $\beta = -.01\pm.05$, $p=.73$), having COVID-like symptoms (first principal component: $\beta = .01\pm.01$, $p=.39$, second principal component: $\beta = .004\pm.009$, $p=.66$), or being close to patients who

tested positive (first principal component: $\beta = -.54 \pm .40$, $p = .17$, second principal component: $\beta = .09 \pm .40$, $p = .81$) modulated the projection to any of the two principal components. This indicates that the observed country-level differences are probably driven by concerns about the societal impact of the pandemic rather than the personal proximity to the virus.

We also reasoned that differences in general personality traits could also be associated to variability on how individuals make moral judgements. We asked participants to complete the Big Five Inventory (32) (see **Methods**) and investigated whether variations across these five dimensions of personality could explain residual variance in the data after controlling for the already described contextual and moral effects. We found that both principal components correlated positively with conscientiousness (first PC: $\beta = .03 \pm .01$, $p = .002$; second PC: $\beta = .03 \pm .01$, $p = .01$) and neuroticism (first PC: $\beta = .04 \pm .01$, $p = .003$; second PC: $\beta = .05 \pm .01$, $p < 10^{-4}$). This suggests that the tendency to prioritize public health over other aspects of wellbeing (either by focusing on human life expectancy or equitable public health) is especially strong in individuals who are diligent and responsible, but also susceptible to experience negative emotions that are prevalent during the pandemic like anxiety (33) and fear (34). We also found that agreeableness was not associated with any of the two components, a result which might be explained by the fact that this trait strongly correlates with instrumental harm (negative correlation, $r = -.09$, $p < 10^{-10}$) and impartial beneficence (positive correlation, $r = .10$, $p < 10^{-13}$). We again emphasize that these results reflect personality modulations on moral predispositions above and beyond the effects given by individual differences utilitarian scores.

Two personality traits, extraversion and openness, showed different effects on the two principal components, which differ on whether it is acceptable to prioritize younger people over older and animal lives (**Figure 1C**). More extraversion was associated to higher scores in the first principal component and lower scores in the second (first PC: $\beta = .05 \pm .01$, $p < 10^{-4}$; second PC: $\beta = -.06 \pm .01$, $p < 10^{-6}$). This is in line with previous observations that high extraversion is associated with psychopathic tendencies (35) and the propensity to break rules (36).

Openness showed a strong negative correlation with the first, but not the second, principal component (first PC: $\beta = -.08 \pm .01$, $p < 10^{-13}$; second PC: $\beta = -.02 \pm .01$, $p = .05$) which is consistent with research showing that scoring lowly on this trait is associated with the endorsement of different forms of prejudice, including ageism (37) and speciesism (38).

Discussion

The analyses presented in this study suggests that inter-individual differences in moral reasoning play a key role on shaping judgements about problems of the COVID-19 crisis. These effects are robust to the inclusion of a long list of control variables (see **Tables S2 and S3** for details) that index individual differences in contextual, personality, and demographic factors. We believe that these findings provide strong empirical support for a recently proposed bidimensional theory of utilitarian reasoning (8). According to dual-process models, utilitarian decisions result from the inhibition of automatic and emotional deontological responses and the subsequent engagement with a rational cost-benefit analysis (2, 28). Our data, however, provides three points of conflict with this classic unidimensional partition between reason and emotion in moral choices (39, 40). First, we found that responses to moral problems of the pandemic were projected onto a bidimensional space and that each dimension was highly correlated with two theoretically different aspects of utilitarian thinking (instrumental harm and impartial beneficence). Second, we found that scoring highly on impartial beneficence was associated with higher levels of distress while making utilitarian judgements, a result which is at odds with the idea that utilitarian choices result from “cold” cost-benefit calculations (41). Third, the projections onto both principal components were larger in people scoring high in neuroticism, suggesting a role played by emotional sensitivity in making utilitarian judgements. Altogether, these findings suggest that the classic *emotional-versus-rational* partition of moral choice may oversimplify people’s reactions to complex issues related to the pandemic.

In Study 1, we generated a large-scale dataset of moral preferences in Latin America, a region that was seriously impacted by the pandemic, and that had contributed to more than

30% of the COVID-19 confirmed deaths by the time the study was performed. Our findings suggest that the crisis may have shifted the focus of utilitarian thinking (generally concerned about several dimensions of wellbeing) to setting a clear priority on public health, given that participants from countries where the pandemic had a more severe impact had a higher projection onto the two principal components. While the evidence presented here is correlational, this study represents a first step towards understanding country-level differences in policy preferences while controlling for individual-level variations in moral and personality variables, among others.

In Study 2, we showed that the utilitarian prioritization of public health over other aspects of wellbeing is a robust empirical observation, despite this replication study used a different sampling strategy and was performed in a different country, language, and moment of the pandemic. While these data show that our main results are replicable in a very different setting, further research is needed to assert whether the utilitarian focus on public health is currently present in cross-cultural samples including other regions of the World (42, 43). In this sense, we believe that the methodology described here can be scaled up to performing large-scale online behavioral studies based on big data obtained from social media (44, 45).

In conclusion, this research organized people's judgements about contemporary problems of the pandemic according to well-established moral principles. This classification should be an important input for policymakers aiming at constructing policies that represent their citizens' preferences. On top of that, we believe that understanding people's moral fingerprints about these issues can help political leaders to design more persuasive messages and improve the communication of public policies to address the crisis (46, 47).

Methods

Context

Study 1 was part of an initiative called “*TEDxperiments*” aimed at constructing knowledge on human cognition by performing behavioral experiments with large audiences (www.tedxriodelaplata.org/tedxperiments). Previous editions studied the use of a competition bias in a ‘zero-sum fallacy’ game (23), the role of deliberation in the wisdom of crowds (26), and the factors underlying consensus in polarized moral debates (27). The study was in collaboration with “*El Gato y la Caja*”, a science popularization project that has previously performed large-scale behavioral studies on different aspects of human psychology (24, 25). In this occasion, the invitation to participate in the study was distributed online through the social media accounts of *TEDxperiments* and *El Gato y la Caja*. The procedure was approved by the ethics committee of CEMIC (Centro de Educación Médica e Investigaciones Clínicas Norberto Quirno).

Participants

The invitation to perform Study 1 was first published on 6th May 2020 and we collected data until 31st May 2020. During this time window, we recruited $N = 15,420$ participants (60.1% female, mean age: 32.0 y.o., range [18-95] years), most of them proceeding from Argentina ($n = 14,443$). The remaining 977 participants were from Mexico ($n = 214$), Uruguay ($n = 204$), Peru ($n = 119$), Ecuador ($n = 107$), Bolivia ($n = 106$), Colombia ($n = 99$), Chile ($n = 92$), Honduras ($n = 37$), and Panama ($n = 32$).

Procedure

Participants read five scenarios related to the COVID-19 crisis in randomized order, as well as the trolley problem in either its personal or impersonal version. Half of the participants were presented with the trolley problem before reading the five scenarios about the pandemic, and

the other half after. Each scenario described a clear action and we asked participants to report the acceptability of that action (in a scale from “completely wrong” to “completely right”) and their confidence about the provided answer (from “I feel completely unsure” to “I feel completely sure”). We also asked them to imagine that they were the person responsible for deciding what to do in that situation and report the amount of distress that they think they would feel (from “none” to “a great deal”). In all cases, participants input their answers by moving three sliders on the screen, providing us a value that ranged from 0 to 100. After reading the five scenarios about the pandemic and the trolley problem, participants completed the Oxford Utilitarianism Scale. Then, they were asked to report their age, gender, and whether they were diagnosed with COVID-19 and/or knew someone who did. Finally, they were invited to complete a standard personality test (see below for details).

Scenarios

The first scenario in **Figure 1A** asks whether it is acceptable to allow the government to collect sensitive private data to allow tracking the virus. It reads: *“Suppose we can go back in time before the pandemic had exploded world-wide. A new technology allows governments to automatically collect personal information about our movements and the people we have physical contact with. Governments can also have access to biometric data such as our temperature and respiratory frequency. This allows identifying people at risk of having COVID-19 and limiting the propagation of the virus by selecting who is quarantined and who is not. The government of the country where you live evaluates a plan to use this technology but understands that it does not protect the personal data of its citizens and decides not to approve it. Not approving this plan is...”*.

The second scenario is about forbidding public gatherings, including several business activities, until a vaccine is found. It reads: *“The near future of our societies and how we will get out of the periods of strict confinement is unclear, among other things because there are great unknowns in the mutation rates of the virus, its seasonal variance, the probability of*

being re-infected, and the degree of immunity that we will reach. Societies are likely to create a compromise between minimizing health risks and preserving our desired lifestyles. Knowing this, a government decides to forbid all kinds of public gatherings (including restaurants, bars, concerts, cinemas, and theatres) until a vaccine is found. The government's decision is...". As a clarification, we stress that in this scenario assumes that the development of a vaccine is equivalent to the end of the COVID-19 crisis. However, to date, there is no evidence on whether vaccines will be sufficiently effectively to overcome the life-destroying effects of social activities. The aim of this scenario is to ask participants whether at present situation (no vaccine) it is correct to forbid all public gatherings or not. This has been part of a very active public debate with very different attitudes and responses in different societies.

The third scenario deals with the tension between notifying a COVID-19 protocol breach versus protecting a close friend from facing prison. It reads: *"One of your close friends has high fever and was recently tested for COVID-19. The health authorities of your country have dictated that he should remain at home, in complete isolation, until the results are back. During this period, your friend went once to the supermarket to buy food. If the authorities find out about this, your friend might face prison and have a criminal record. You still do not know whether he will test positive for COVID-19 but decided to call the police and tell the authorities about this situation. Telling the authorities about this situation is..."*

The fourth scenario is about the assignment of scarce medical resources and asks whether younger people should be prioritized over older patients. It reads: *"Ventilators have become in many countries a limited resource and a medical bottleneck. An ethics committee is deciding how to set priorities on who will be treated and who will not. One option is to treat people on a first-come first-serve basis. A second argument is made that significantly younger patients (under 30 years old) should be prioritized and that, even when chances of recovery are considered to be the same, the younger patient should be treated before older patients. The ethics committee opts to implement the second option. This decision is..."*

The fifth scenario poses a tension between animal rights and accelerating the development of a vaccine for COVID-19. It reads: *“A laboratory has all the necessary resources to develop a vaccine against coronavirus in only weeks. However, to do so, the researchers need to violate ethical guidelines for the use of animals in scientific research. Particularly, the experiments will cause excessive stress and pain to hundreds of rodents. The laboratory requests an exceptional permission to perform these experiments, but the government decides to reject this request. Rejecting this permission is...”*

For the purposes of data analysis, we reverse-coded the acceptability ratings to the first and fifth dilemma, so that positive values indicate a prioritization to public health and a prioritization to saving younger human lives (**Figure 1A**).

Trolley Problems

The impersonal version of the trolley problem was presented as follows: *“A runaway train is speeding down the tracks towards five people who will be killed if the train continues on its present course. You are standing next to the tracks, but you are too far away to warn them. Next to you there is a control switch that can redirect the train onto a different track, where there is only one person. If you decide to flip the control switch, that will provoke the death of one person, but the train will not continue its course and you will have saved the lives of five people. Flipping the control switch is ...”*.

The personal variant of the same classic scenario read: *“A runaway train is speeding down the tracks towards five people who will be killed if the train continues on its present course. You are standing next to the tracks, but you are too far away to warn them. Next to you there is someone else and, if this person falls to the tracks, the train will derail. If you decide to push this person onto the tracks, that will provoke the death of this person, but the train will not continue its course and you will have saved the lives of five people. Pushing the person onto the tracks is...”*

Utilitarianism Scale

Individual differences in utilitarian tendencies were measured using the Oxford Utilitarianism Scale (OUS), which was developed and validated in a previous study (8). The scale consists in 9 items presented in randomized order, 4 of which measure permissive attitudes towards instrumental harm (OUS-IH, instrumental harm) and the remaining 5 evaluate people's impartial concern for the greater good (OUS-IB, impartial beneficence). All 9 items were read on the same page and responses were collected using a 7-point scale from "strongly disagree" to "strongly agree".

The four items of the OUS-IH sub-scale are: 1) *"It is morally right to harm an innocent person if harming them is a necessary means to helping several other innocent people"*; 2) *"If the only way to ensure the overall well-being and happiness of the people is through the use of political oppression for a short, limited period, then political oppression should be used"*; 3) *"It is permissible to torture an innocent person if this would be necessary to provide information to prevent a bomb going off that would kill hundreds of people"*; 4) *"Sometimes it is morally necessary for innocent people to die as collateral damage—if more people are saved overall"*.

The five items of the OUS-IB sub-scale are: 1) *"If the only way to save another person's life during an emergency is to sacrifice one's own leg, then one is morally required to make this sacrifice"*; 2) *"From a moral point of view, we should feel obliged to give one of our kidneys to a person with kidney failure since we don't need two kidneys to survive, but really only one to be healthy"*; 3) *"From a moral perspective, people should care about the well-being of all human beings on the planet equally; they should not favour the well-being of people who are especially close to them either physically or emotionally"*; 4) *"It is just as wrong to fail to help someone as it is to actively harm them yourself"*; and 5) *"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"*.

Principal Component Analysis (PCA)

We performed a PCA on the acceptability ratings across the five scenarios related to the COVID-19 pandemic. To identify the number of principal components that explained a significantly large proportion of the variance, we performed a random permutation analysis. We generated 10,000 surrogated datasets by randomly shuffling the participants' labels for each scenario. Applying PCA on each surrogate, we estimated the distribution of explained variances for the first to fifth component under the null hypothesis that there are no correlations in the data (grey shade in **Figure 3A**). We then estimated the p-value of this statistical test as the fraction of times that a given component in the simulated data accounted for more variance than the one observed explained by that same component in the real data. We found that the first and second principal components explained more variance than any of the first and second components in the surrogated dataset, whereas the remaining three components explained less variance than what expected by pure chance.

Pre-registered replication study

Study 2 aimed at providing a high-powered pre-registered replication of Study 1 while using a sample which is representative of the United States population. To determine sample size, we performed a Monte Carlo power analysis based on the data collected in the original study. For each simulated sample size, we randomly sub-sampled the dataset 10,000 times without replacement and repeated the analyses performed in Study 1. We then estimated power as the fraction of times that we were able to successfully replicate three observations: i) the existence of two principal components explaining a significant proportion of the variance in the data; ii) the fact that these two dimensions would reflect human life expectancy and equitable public health (**Figure 1C**); and iii) the association between each principal component with a instrumental harm and impartial beneficence. Based on the results of this power analysis (**Figure S3**), we decided to collect data from 1300 participants.

On 23rd October 2020, we pre-registered the hypotheses, design, and analysis plan of Study 2 (<https://aspredicted.org/59kk7.pdf>). Data collection started on 23 October 2020 and proceeded until 30 October 2020. In this time window, we recruited the target sample size (aged 18-89, 48.3% female). This sample was obtained through Prolific, an online platform to recruit human participants for scientific research (<https://www.prolific.co/>). To obtain this sample, participants were stratified across three demographics: age, sex, and ethnicity. Recruited participants were divided into subgroups that have similar proportions to the national population, based on data obtained from the US Census Bureau (<https://www.census.gov/>). The demographic characteristics of our sample and their comparison with the United States population are available in **Table S1**.

Country-level COVID-19 data

Data about COVID-19 confirmed deaths and cases was obtained through a free and open-access platform called “Our World in Data” that publishes research and data on some of the world’s largest problems including health, poverty, and human rights, among many others (<https://ourworldindata.org/>). Data about coronavirus published on that platform rely on figures issued by official authorities such as governments and ministries of health. The map displayed in **Figure 4A** was first obtained through the website of “Our World in Data” in .svg format and adapted to reflect custom data ranges.

Other individual-level contextual factors

In principle, one may reasonably ask whether showing COVID-like symptoms is a better predictor of moral responses than being diagnosed with COVID-19. This question is relevant given that a large fraction of carriers of COVID-19 are known to be asymptomatic. However, in our sample the proportion of diagnosed individuals who were entirely asymptomatic is expected to be very low, given that the criterion to get tested for COVID-19 in most Latin American countries is showing COVID-like symptoms. In Study 1, we asked participants to

indicate whether they had fever. If they were diagnosed with COVID-19, we asked if that symptom showed up at some point during the course of the disease. If they had not been diagnosed with COVID-19, we asked if they have had fever in the previous week. We observed that 94.7% of the participants who were diagnosed with COVID-19 also indicated to have had fever during the course of the disease. We also observed that 1.7% of our sample indicated having fever but not being diagnosed with COVID-19. To study whether symptoms could predict moral responses we added this binary variable as a predictor in the regression estimated in **Tables S2 and S3**. However, we found that it did not modulate any of the two first principal components (first principal component: $\beta = .01 \pm .01$, $p = .39$; second principal component: $\beta = .004 \pm .009$, $p = .66$). This finding is consistent with our previous observation that individual-level contextual variables did not modulate responses to the moral scenarios (**Tables S2 and S3**).

Individual and contextual effects on self-reported distress about COVID-19 problems

We observed that individuals scoring high on instrumental harm reported less distress about the trolley dilemma and the opposite effect was found for people with high scores on impartial beneficence (**Figure S2**). One question that arises from these findings is whether this effect hold for COVID-19 problems. To answer this question, we looked at the correlation between each sub-score of the Oxford Utilitarianism Scale and the average rating of self-reported distress across the five moral dilemmas about the pandemic. In accordance with our observations based on the Trolley Problem (**Figure S2**), we found that distress negatively correlated with instrumental harm ($r = -.07$, $p = 4 \times 10^{-16}$) and positively correlated with impartial beneficence ($r = .15$, $p = 3 \times 10^{-73}$). We also asked whether contextual variables modulated self-reported measures of distress and found that the only significant predictor of less distress when considering moral problems of the pandemic was being diagnosed with COVID-19 ($\beta = -.47 \pm .23$, $p = .03$).

Big Five personality traits

In Study 1, we assessed personality using the 44-item Big Five Inventory (32), measuring participants on five dimensions of personality (extraversion, agreeableness, conscientiousness, neuroticism, openness). Each item is presented in a 5-choice answer format that ranges from complete disagreement (1 = very false for me) to complete agreement (5 = very true for me). We used a version in *Rioplatense* Spanish, adapted and validated in previous research (48).

Multivariate regression

To measure how the projection onto the first two principal components depended on moral, contextual, personality, and demographic variables, we performed a multivariate linear regression. This regression included a total of 15 predictor variables, as displayed in **Tables S2 and S3**. Within those 15 regressors, we included two binary variables related to the structure of the study (i.e., “experimental” variables): the order of presentation of the trolley problem (0: before the five main scenarios, 1: after reading the five main scenarios) and its version (0: impersonal, 1: personal).

Data availability

All data and codes to reproduce our findings are available at <https://osf.io/mwe9r/>

References

1. Foot P (1967) The problem of abortion and the doctrine of double effect. *Oxford Review* 5: 5-15.
2. Greene JD, Sommerville RB, Nystrom LE, Darley JM & Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537): 2105-2108.
3. Skitka LJ, Bauman CW & Sargis EG (2005) Moral conviction: Another contributor to attitude strength or something more?. *J Pers Soc Psychol* 88(6): 895.
4. Crockett MJ, Clark L, Hauser MD & Robbins TW (2010) Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci U S A* 107(40): 17433-17438.
5. Costa A, et al (2014) Your morals depend on language. *Plos One* 9(4): e94842.
6. Christensen JF, Flexas A, Calabrese M, Gut NK & Gomila A (2014) Moral judgment reloaded: A moral dilemma validation study. *Frontiers in Psychology* 5: 607.
7. Parnaments P, et al (2015) Biasing moral decisions by exploiting the dynamics of eye gaze. *Proc Natl Acad Sci U S A* 112(13): 4170-4175.
8. Kahane G, et al (2018) Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychol Rev* 125(2): 131.
9. Teper R, Inzlicht M & Page-Gould E (2011) Are we more moral than we think? exploring the role of affect in moral behavior and moral forecasting. *Psychological Science* 22(4): 553-558.
10. FeldmanHall O, et al (2012) What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition* 123(3): 434-441.
11. Camerer C & Mobbs D (2017) Differences in behavior and brain activity during hypothetical and real choices. *Trends Cogn Sci (Regul Ed)* 21(1): 46-56.
12. Gonzalez-Gadea ML, Ibanez A & Sigman M (2018) Schadenfreude is higher in real-life situations compared to hypothetical scenarios. *PloS One* 13(10)
13. Loewenstein G (2005) Hot-cold empathy gaps and medical decision making. *Health Psychology* 24(4S): S49.
14. Ariely D & Loewenstein G (2006) The heat of the moment: The effect of sexual arousal on sexual decision making. *J Behav Decis Making* 19(2): 87-98.
15. Nordgren LF, Banas K & MacDonald G (2011) Empathy gaps for social pain: Why people underestimate the pain of social suffering. *J Pers Soc Psychol* 100(1): 120.
16. Kang MJ & Camerer CF (2013) fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Frontiers in Neuroscience* 7: 104.
17. Emanuel EJ, et al (2020) Fair allocation of scarce medical resources in the time of covid-19. *The New England Journal of Medicine* 382
18. White DB & Lo B (2020) A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA* 323(18): 1773-1774.
19. Ienca M & Vayena E (2020) On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med* 26(4): 463-464.
20. Atkeson A (2020) What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. *NBER Working Paper* 26867
21. Awad E, et al (2018) The moral machine experiment. *Nature* 563(7729): 59-64.
22. Bentham J (1983) The collected works of jeremy bentham: Deontology. together with a table of the springs of action and the article on utilitarianism
23. Niella T, Stier-Moses N & Sigman M (2016) Nudging cooperation in a crowd experiment. *Plos One* 11(1): e0147125.
24. Zimmerman F, et al (2016) Arithmetic on your phone: A large scale investigation of simple additions and multiplications. *Plos One* 11(12): e0168431.
25. Rieznik A, et al (2017) A massive experiment on choice blindness in political decisions: Confidence, confabulation, and unconscious detection of self-deception. *Plos One* 12(2): e0171108.

26. Navajas J, Niella T, Garbulsky G, Bahrami B & Sigman M (2018) Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour* 2: 126-132.
27. Navajas J, *et al* (2019) Reaching consensus in polarized moral debates. *Current Biology* 29(23): 4124-4129. e6.
28. Greene JD (2007) Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends Cogn Sci (Regul Ed)* 11(8): 322-323.
29. Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58(3): 453-467.
30. Dollinger SJ & LaMartina TK (1998) A note on moral reasoning and the five-factor model. *Journal of Social Behavior and Personality* 13(2): 349.
31. Hirsh JB, DeYoung CG, Xu X & Peterson JB (2010) Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Person Soc Psychol Bull* 36(5): 655-664.
32. John OP (1990) in *Handbook of personality: Theory and Research*. ed Pervin L (Guilford Press, New York),
33. Qiu J, *et al* (2020) A nationwide survey of psychological distress among chinese people in the COVID-19 epidemic: Implications and policy recommendations. *Gen Psychiatr* 33(2): e100213-2020-100213. eCollection 2020.
34. Harper CA, Satchell LP, Fido D & Latzman RD (2020) Functional fear predicts public health compliance in the COVID-19 pandemic. *International Journal of Mental Health and Addiction*
35. Newman JP, Widom CS & Nathan S (1985) Passive avoidance in syndromes of disinhibition: Psychopathy and extraversion. *J Pers Soc Psychol* 48(5): 1316.
36. Rushton JP & Chrisjohn RD (1981) Extraversion, neuroticism, psychoticism and self-reported delinquency: Evidence from eight separate samples. *Personality and Individual Differences* 2(1): 11-20.
37. Allan LJ, Johnson JA & Emerson SD (2014) The role of individual difference variables in ageism. *Personality and Individual Differences* 59: 32-37.
38. Caviola L, Everett JA & Faber NS (2019) The moral standing of animals: Towards a psychology of speciesism. *J Pers Soc Psychol* 116(6): 1011.
39. Mercier H (2011) What good is moral reasoning?. *Mind & Society* 10(2): 131-148.
40. Heinzelmann N (2018) Deontology defended. *Synthese* 195(12): 5197-5216.
41. Greene JD, Nystrom LE, Engell AD, Darley JM & Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44(2): 389-400.
42. Sznycer D, *et al* (2017) Cross-cultural regularities in the cognitive architecture of pride. *Proc Natl Acad Sci U S A* 114(8): 1874-1879.
43. Sznycer D, *et al* (2018) Cross-cultural invariances in the architecture of shame. *Proc Natl Acad Sci U S A* 115(39): 9702-9707.
44. Zhou Z & Makse HA (2019) Artificial intelligence for elections: The case of 2019 argentina primary and presidential election. *ArXiv Preprint*
45. Bovet A & Makse HA (2019) Influence of fake news in twitter during the 2016 US presidential election. *Nature Communications* 10(1): 1-14.
46. Everett JA, Colombatto C, Chituc V, Brady WJ & Crockett M (2020) The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic
47. Bilancini E, Boncinelli L, Capraro V, Celadin T & Di Paolo R (2020) The effect of norm-based messages on reading and understanding COVID-19 pandemic response governmental rules. *ArXiv Preprint arXiv:2005.03998*
48. Solano C (2005) Técnicas de evaluación psicológica para ámbitos militares. *Buenos Aires.Argentina: Paidós*.

Acknowledgements

We thank Victoria Milano, Rocco Di Tella, Juan Ignacio Cuile, Maria Agustina Nahas, Laura Gonzalez, Juan Manuel Garrido, Pablo Gonzalez, Nuria Caceres, Marco Sartorio, and Lucia Freira for assistance with data collection.

Funding

J.N. and M.S. were supported by the James McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition—Scholar Award (Grant #220020334), and by Agencia Nacional de Promoción Científica y Tecnológica (Argentina)—Préstamo BID PICT (Grant #2013-1653).

Supplementary Material

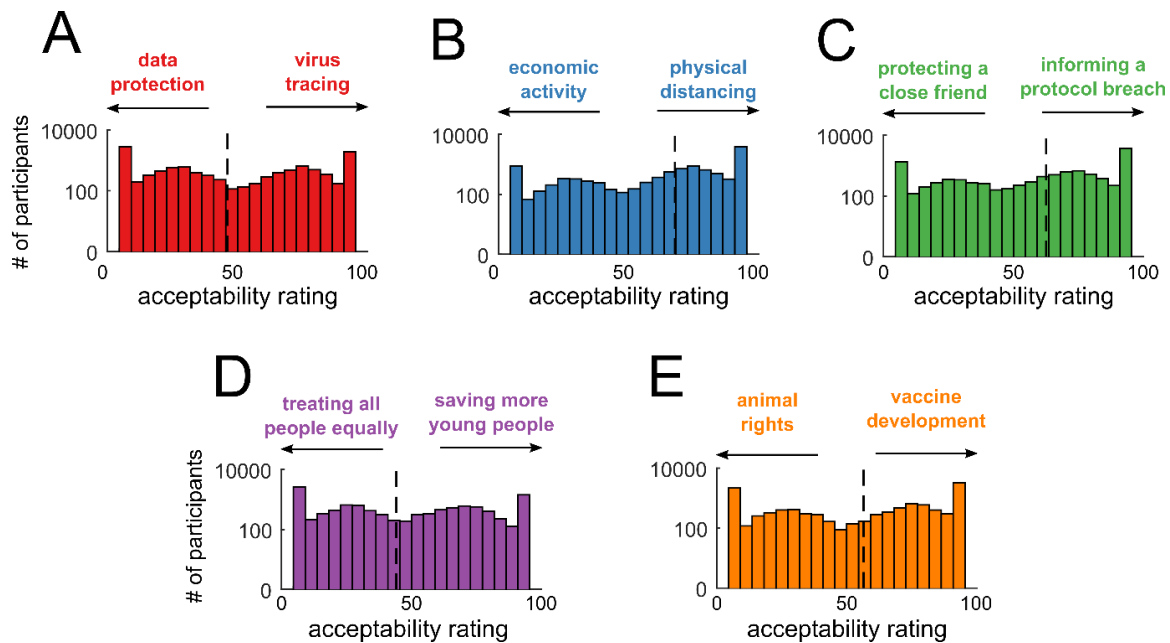


Fig. S1 Histograms of acceptability ratings across five moral dilemmas related to the COVID-19 pandemic. We developed five moral scenarios describing a clear action. Participants read them in randomized order and provided acceptability ratings (see Materials and Methods for details). Three of these scenarios pose a tension between prioritizing public health versus other aspects of wellbeing (A, B, and C). The remaining two scenarios (D and E) ask whether it is acceptable to prioritize young or human lives over older and animal lives. To appreciate the diversity of responses, histograms are displayed using a log-scale in the y-axis.

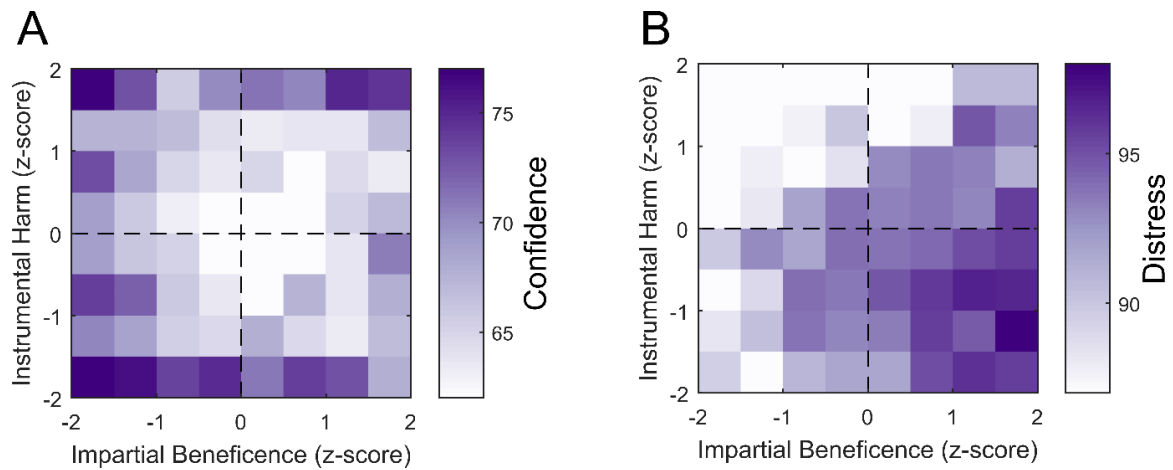


Fig. S2. Confidence judgements and self-reported distress about the trolley problem. Images show average values of confidence (A) and distress (B) as a function of participants' impartial beneficence (OUS-IB, x-axis) and instrumental harm (OUS-IH, y-axis). A) We observed that confidence depended on the extremity of each sub-scales suggesting that low or high scores on any of these two dimensions was associated with higher confidence about moral judgements (bi-variate regression of absolute z-scored OUS sub-scales on confidence in the trolley problem, effect of instrumental harm: $\beta = .22 \pm .02$, $p < 10^{-28}$, effect of impartial beneficence: $\beta = .13 \pm .02$, $p < 10^{-11}$). B) Self-reported values of distress when considering the trolley problem showed opposing modulations by each of the sub-scales. While higher values of impartial beneficence correlated with higher distress, higher values of instrumental harm were associated with lower distress (bi-variate regression of z-scored sub-scales on distress about the trolley problem, effect of instrumental harm: $\beta = -.15 \pm .01$, $p < 10^{-41}$, effect of impartial beneficence: $\beta = .14 \pm .01$, $p < 10^{-35}$).

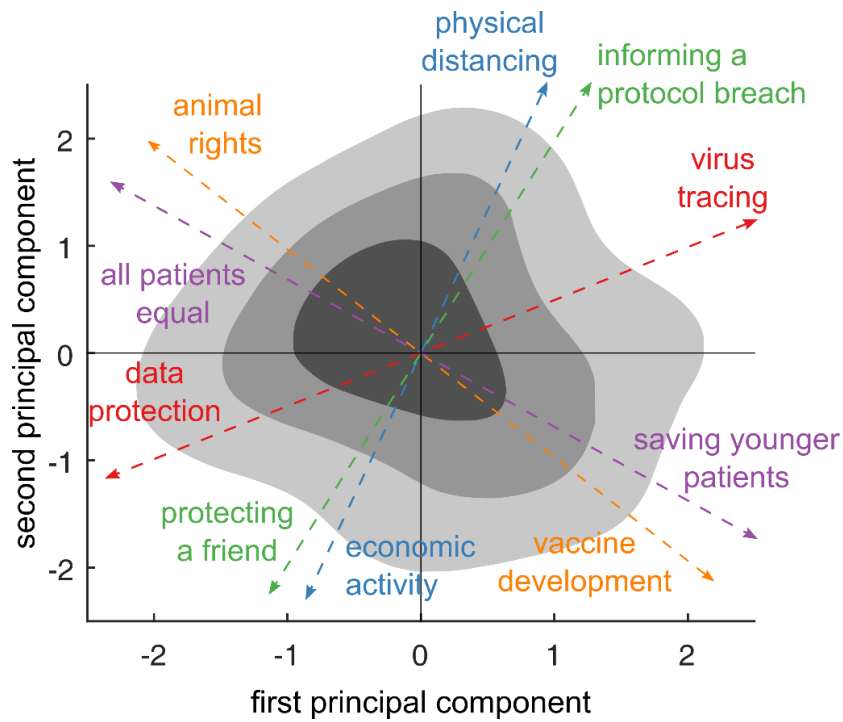


Fig. S3. PCA bi-plot of moral responses to the scenarios about the COVID-19 pandemic.. The two axes represent the two first principal components in the data. The rotated dilemma-by-dilemma axes are obtained by projecting the loadings of each dilemma onto these two components. For example, the first three dilemmas have positive loadings in the first two principal components and so they go from the bottom-left quadrant to the upper-right quadrant of the bi-plot. The last two dilemmas have positive loadings in the first component but negative loadings in the second component and so they go from the upper-left quadrant to the lower-right quadrant of the bi-plot. Grey shades show the joint probability distribution on this space with darker colors indicating higher density.

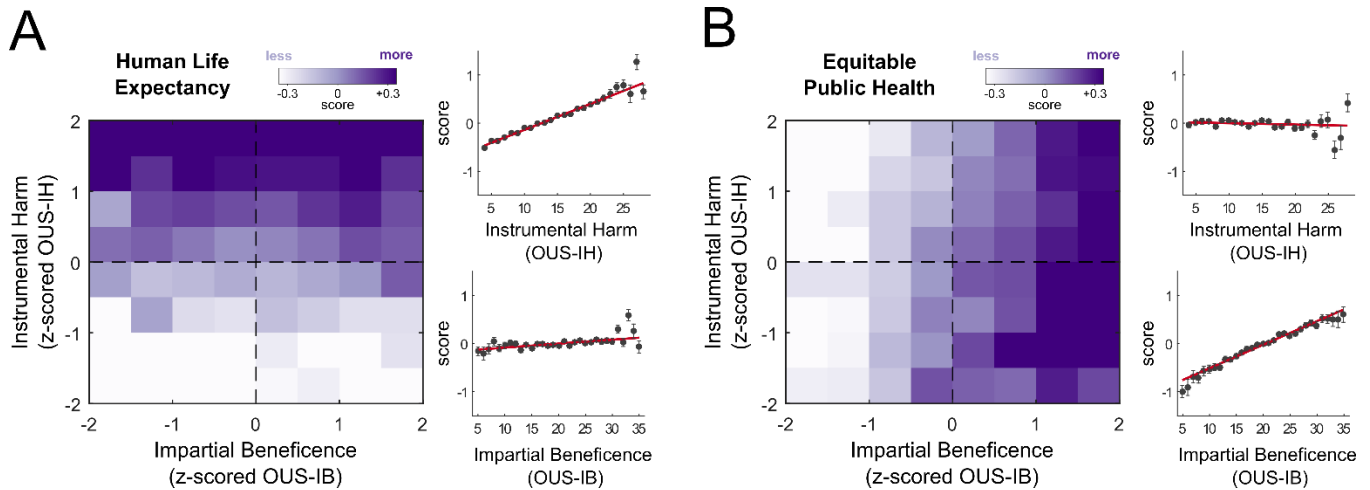


Figure S4. Correlations between projections onto the first two principal components and individual differences in utilitarian reasoning styles. A) Using the data of Study 1, we show the relationship between different dimensions of utilitarian reasoning and the first principal component (i.e., human life expectancy). Images show the average score of each component for participants with a given value of z-scored impartial beneficence (OUS-IB, x-axis) and instrumental harm (OUS-IH, y-axis). This image is similar to the one displayed in Figure 3C but with higher resolution. Insets show bin-scatter plots with black dots representing average PCA scores for a given value of OUS-IH (upper panel) and OUS-IB (lower panel). Best fitting lines are display in red. D) Same as C but for the second principal component, which in Study 1 reflects a concern for equitable public health.

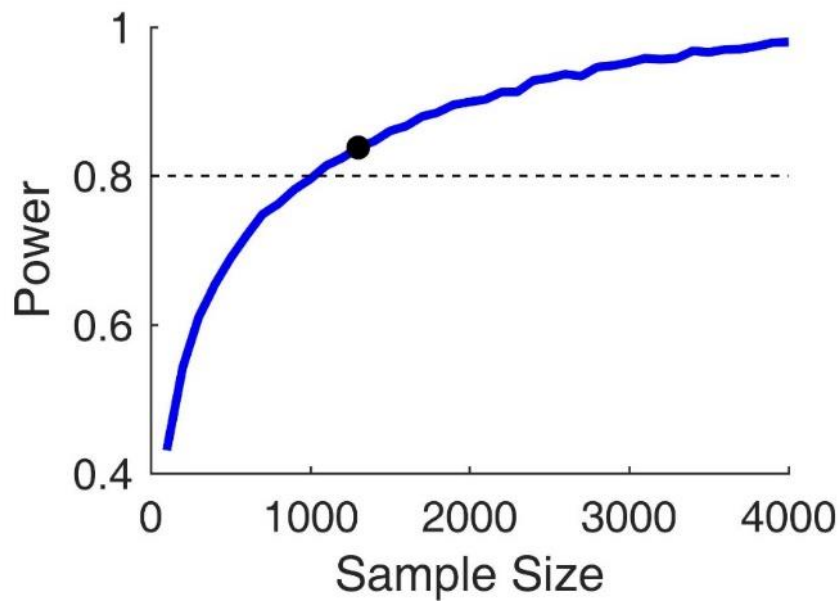


Figure S5. Power analysis used to determine sample size in Study 2. We performed a Monte Carlo power analysis based on the data collected in Study 1. For each simulated sample size, we randomly sub-sampled our dataset 10,000 times without replacement and repeated the same analysis performed in Study 1. We then estimated power as the fraction of times that we were able to successfully replicate the three main results of Study 1: 1) the existence of two principal components explaining a significant proportion of the variance in the data, 2) the fact that these two dimensions reflect Human Life Expectancy and Equitable Public Health, and 3) the positive correlation between these two dimensions with Instrumental Harm and Impartial Beneficence, respectively. Based on this analysis, we decided to collect data from N=1300 participants, which yields an estimated power of 83.8%.

Variable	US Census	Study 2
Sex		
Female	49%	49%
Male	51%	51%
Age		
18-27	17%	17%
28-37	18%	30%
38-47	18%	21%
48-57	16%	13%
58+	29%	19%
Ethnicity		
Asian	8%	7%
Black	15%	13%
Mixed	5%	4%
Other	3%	3%
White	69%	73%

Table S1. Comparison between the demographic characteristics of the United States population and the sample used in Study 2. We collected data from a sample of participants from the United States obtained through Prolific, an online platform to recruit human participants for scientific research (<https://www.prolific.co/>). To obtain this sample, participants were stratified across three demographics: age, sex, and ethnicity. Recruited participants were divided into subgroups that have similar proportions to the national population, based on data obtained from the US Census Bureau (<https://www.census.gov/>).

<i>Type of variable</i>	<i>Variable</i>	<i>Estimate</i>	<i>SE</i>	<i>t value</i>	<i>p value</i>
Moral	Instrumental Harm (IH)	0.31	0.04	8.63	7.48×10^{-18}
	Impartial Beneficence (IB)	0.01	0.03	0.27	0.78
	IB*IH	0.01	0.05	0.17	0.86
Contextual	COVID-19 deaths per million	0.05	0.01	4.01	4.62×10^{-5}
	Tested positive for COVID-19	0.01	0.01	0.81	0.42
	Acquaintance had COVID-19	0.01	0.01	0.94	0.35
Personality	Openness	-0.08	0.01	-7.46	9.45×10^{-14}
	Conscientiousness	0.03	0.01	3.03	0.002
	Extraversion	0.05	0.01	4.41	1.03×10^{-5}
	Agreeableness	0.01	0.01	1.29	0.20
	Neuroticism	0.04	0.01	2.96	0.003
Demographic	Gender (0: male, 1: fem)	-0.03	0.01	-2.70	0.007
	Age	0.08	0.01	6.35	2.33×10^{-10}
Experimental	Version of the Trolley Problem (0: impersonal, 1: personal)	0.02	0.02	1.11	0.27
	Presentation of Trolley Problem (0: before, 1: after)	0.02	0.02	1.09	0.28

Table S2. Multivariate regression on the first principal component (Human Life Expectancy).

Linear regression model with moral, contextual, personality, demographic and experimental variables.

Coefficient estimates are in standardized units. Variables in red (orange) showed a significant modulation with a 1% (5%) significance level.

Type of variable	Variable	Estimate	SE	t value	p value
Moral	Instrumental Harm (IH)	-0.06	0.04	-1.67	0.10
	Impartial Beneficence (IB)	0.25	0.03	9.38	8.64×10 ⁻²¹
	IB*IH	-0.07	0.05	-1.54	0.12
Contextual	COVID-19 deaths per million	0.07	0.01	6.09	1.15×10 ⁻⁹
	Tested positive for COVID-19	-0.01	0.01	-0.62	0.54
	Acquaintance had COVID-19	-0.01	0.01	-0.97	0.33
Personality	Openness	-0.02	0.01	-1.95	0.05
	Conscientiousness	0.03	0.01	2.51	0.01
	Extraversion	-0.06	0.01	-5.17	2.44×10 ⁻⁷
	Agreeableness	0.01	0.01	1.24	0.22
	Neuroticism	0.05	0.01	4.08	4.61×10 ⁻⁵
Demographic	Gender (0: male, 1: fem)	0.14	0.01	12.40	5.94×10 ⁻³⁵
	Age	-0.09	0.01	-6.64	3.37×10 ⁻¹¹
Experimental	Version of the Trolley Problem (0: impersonal, 1: personal)	0.001	0.02	0.03	0.98
	Presentation of Trolley Problem (0: before, 1: after)	-0.03	0.02	-1.36	0.17

Table S3. Multivariate regression on the second principal component (Equitable Public Health).

Linear regression model with moral, contextual, personality, demographic and experimental variables. Coefficient estimates are in standardized units. Variables in red (orange) showed a significant modulation with a 1% (5%) significance level.