Supplementary Information for:

Moral dilemmas and trust in leaders during a global health crisis

Jim A.C. Everett†, Clara Colombatto†, Edmond Awad, Paulo Boggio, Björn Bos, William J. Brady, Megha Chawla, Vladimir Chituc, Dongil Chung, Moritz A. Drupp, Srishti Goel, Brit Grosskopf, Frederik Hjorth, Alissa Ji, Caleb Kealoha, Judy S. Kim, Yangfei Lin, Yina Ma, Michel A. Maréchal, Federico Mancinelli, Christoph Mathys, Asmus L. Olsen, Graeme Pearce, Annayah M.B. Prosser, Niv Reggev, Nicholas Sabin, Julien Senn, Yeon Soon Shin, Walter Sinnott-Armstrong, Hallgeir Sjåstad, Madelijn Strick, Sunhae Sul, Lars Tummers, Monique Turner, Hongbo Yu, Yoonseo Zoh, Molly J. Crockett*

†These authors contributed equally

*Correspondence to: molly.crockett@yale.edu

Pilot Data

We conducted two pilots to establish both proof of concept and the feasibility of our methods (see https://osf.io/m9tpu/).

Pilot 1. We recruited a convenience sample of British participants (N = 100, after exclusions N = 98; N women = 56, N men = 40, N participants with another gender identity = 2; 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 in Main Text) 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 “woman” 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/.

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 task, we recruited a convenience sample of U.S. participants (N = 503, after exclusions N = 469; N women = 239, N men = 212, N participants with another gender identity = 18; 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 women = 59, N men = 37, N participants with another gender identity = 4; 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 employed in the Registered Report (see Analysis Plan), for the self-report 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 “woman” 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/.

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 employed 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 “woman” 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/.

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/). 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[1].

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 treating 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 treating 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 it 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 nor wrong			Absolutely morally right

2. Supplementary Method: Voting Task

Voter Participants

Note: Order of utilitarian and non-utilitarian leaders was 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, 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 Task

Note: Selection of either the utilitarian or non-utilitarian leader was 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 region 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 region is arguing that we should consider lifting 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 region 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 region 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 region 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 region 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 treating 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 Results

1. Supplementary Results: Representativeness

To evaluate the representativeness of our samples across gender categories, we computed the difference between expected (based on population characteristics) and actual proportion of participants who identified as men or women, separately for each country. Most differences were under or equal to 5%, suggesting that our samples were broadly nationally representative for gender in the majority of our sampled countries, with two exceptions: Singapore (women were underrepresented by 6%), and United Arab Emirates (men were underrepresented by 21%).

To evaluate representativeness of our samples across age categories, we computed the difference between expected (based on population characteristics) and actual proportion of participants in each age category, separately for each country. Most differences were less than or equal to 5%, suggesting that our samples were broadly nationally representative for age in the majority of our sampled countries. In six countries, participants in older age categories were underrepresented: Kingdom of Saudi Arabia (with an 11% difference between expected and actual proportion of participants in the 45 to 54 age category, 7% in the 55 to 64 age category, and 6% in the over 65 age category – while younger participants were overrepresented by 15% in the 25 to 34 age category), Singapore (6% in the 45 to 54 age category, 12% in the 55 to 64 age category, and 15% in the over 65 age category – while younger participants were overrepresented by 17% in the 25 to 34 age category, and by 11% in the 35 to 44 age category), South Korea (7% in the over 65 age category), United Arab Emirates (8% in the 45 to 54 age category, and 6% in the 55 to 64 age category – while younger participants were overrepresented by 9% in the 18 to 24 age category, by 7% in the 25 to 34 age category), United Kingdom (13% in the over 65 age category – while participants in the 55-64 age category were overrepresented by 9%), and United States (8% in the over 65 age category). In contrast, participants in the over 65 age category were overrepresented in Germany (by 6%).

2. Supplementary Results: Alternative Model Structure for Voting Task

Previous work arguing that linear models should be preferred over logit models is based on tests of models without random effects[2]. Therefore, we first looked at whether the discrepancy between the binomial and linear models was due to an overly complex random effects structure. Indeed, when removing the random effects (of countries and dilemmas), and retaining solely the fixed effects (of demographic variables, own policy preference, and moral dimension), the binomial and linear models converged and both supported our predictions. We found a strong effect of moral dimension on leader choice (main effect of dimension type in binomial model: $B = 1.34$, SE = 0.05, $z = 26.38$, $p < .001$, CI = [1.22, 1.45], OR = 3.80; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.21, SE = 0.02, CI = [0.17, 0.25]; probability of choosing utilitarian leader in Impartial Beneficence dilemmas = 0.50, SE = 0.02, CI =

[0.45, 0.56]; main effect of dimension type in linear model: $B = 0.19$, SE = 0.01, $t = 28.05$, $p < .001$, CI = [0.17, 0.20]; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.30, SE = 0.01, CI = [0.27, 0.33], in Impartial Beneficence dilemmas = 0.49, SE = 0.01, CI = [0.46, 0.52]). Further evidence that the unusual results from the linear pre-registered model were driven by error estimation from an overly complex random effects structure came from a second model where we added countries as random effects with robust standard errors. Here, the model yielded highly significant results ($B = 0.19$, SE = 0.01, $t = 15.78$, $p < .001$, CI = [0.16, 0.21]; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.30, SE = 0.02, CI = [0.26, 0.35], in Impartial Beneficence dilemmas = 0.49, SE = 0.02, CI = [0.43, 0.54]).

Because the random effects structure appeared to cause issues with the linear model, but dilemmas and countries are of clear theoretical interest, we next ran additional exploratory models retaining countries as fixed effects (along with the fixed effects of demographic variables, own policy support, and moral dimension). These analyses again confirmed a strong effect of moral dimension on leader choice, both in the binomial and linear models (main effect of dimension type in binomial model: $B = 1.34$, SE = 0.05, $z = 26.37$, $p < .001$, CI = [1.23, 1.46], OR = 3.84; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.21, SE = 0.02, CI = [0.18, 0.25]; probability of choosing utilitarian leader in Impartial Beneficence dilemmas = 0.51, SE = 0.02, CI = [0.45, 0.57]; main effect of dimension type in linear model: $B = 0.19$, SE = 0.01, $t = 28.08$, $p < .001$, CI = [0.17, 0.20]; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.31, SE = 0.01, CI = [0.27, 0.34], in Impartial Beneficence dilemmas = 0.49, SE = 0.01, CI = [0.46, 0.52]).

Next, we considered models that varied in the complexity of the covariate structure, based on recent recommendations that simpler models should be preferred over those with a complex covariate structure[3,4]. We ran additional exploratory models removing all covariates and simply predicting leader choice from moral dimension. Again, the binomial and linear model both confirmed a strong effect of moral dimension on leader choice (main effect of dimension type in binomial model: $B = 1.37$, SE = 0.04, $z = 35.90$, $p < .001$, CI = [1.28, 1.45], OR = 3.92; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.28, SE = 0.01, CI = [0.26, 0.29]; probability of choosing utilitarian leader in Impartial Beneficence dilemmas = 0.60, SE = 0.01, CI = [0.58, 0.61]; main effect of dimension type in linear model: $B = 0.32$, SE = 0.01, $t = 38.76$, $p < .001$, CI = [0.30, 0.34]; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.28, SE = 0.01, CI = [0.26, 0.29], in Impartial Beneficence dilemmas = 0.60, SE = 0.01, CI = [0.58, 0.61]).

Finally, we employed a disciplined approach to covariate selection[5] by using a double lasso procedure: we began with all demographics (namely gender, age, education, subjective SES, political ideology, and religiosity), own policy support, and country, and identified via two separate lasso regressions the ones that strongly predicted either voting preferences or dimension. Next, we ran a linear model with these empirically supported covariates (all except for education); this again revealed a strong effect of moral dimension on leader choice ($B = 0.19$, SE = 0.01, $t = 28.08$, $p < .001$, CI

= [0.17, 0.20]; probability of choosing utilitarian leader in Instrumental Harm dilemmas = 0.30, SE = 0.01, CI = [0.27, 0.34], in Impartial Beneficence dilemmas = 0.49, SE = 0.01, CI = [0.46, 0.52]). Finally, we verified that the random structure of our data did not alter our results by conducting randomization inference, wherein the effect size obtained in the data (where treatment assignment had been randomized within each country) is compared to a sharp null distribution (where all possible treatment assignments are simulated, and the treatment effect is null for all subjects). This analysis confirmed that the observed effect of dimension on voting choices in the behavioral task (B = 0.19 in a model predicting voting choices from moral dimension along with all demographic covariates and own policy preferences) was significantly different from the null distribution of all possible random assignments ($p < .001$).

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 maximizing aggregate welfare and respective rights and duties.

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 typically only been conducted in a limited number of Western populations (e.g. the US, Belgium 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 younger over older people 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 younger over older people 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-report 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 behavioral) 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). 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.

Supplementary Tables

Supplementary Table 1. Countries, recruitment platforms, survey languages, planned sample size, and final obtained sample sizes. Participants in all countries were 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	Final N
Australia	Lucid	English	1000	994
Brasil	Lucid	Portguese	1000	1298
Canada	Lucid	English	1000	1102
Chile	CESS Santiago	Spanish	1000	1468
China	Lucid	Chinese	1000	1517
Denmark	Epinion	Danish	1000	1155
France	Lucid	French	1000	1073
Germany	Lucid	German	1000	1192
India	Lucid	Hindi	1000	1269
Israel	Panel HaMidgam	Hebrew	1000	851
Italy	Lucid	Italian	1000	1132
Kingdom of Saudi Arabia	Lucid	Arabic	500	757
Mexico	Lucid	Spanish	1000	1260
Netherlands	Lucid	Dutch	1000	1143
Norway	Norstat	Norweigan	1000	1217
Singapore	Lucid	English	1000	848
South Africa	Lucid	English	1000	1120
South Korea	Dataspring	Korean	1000	797

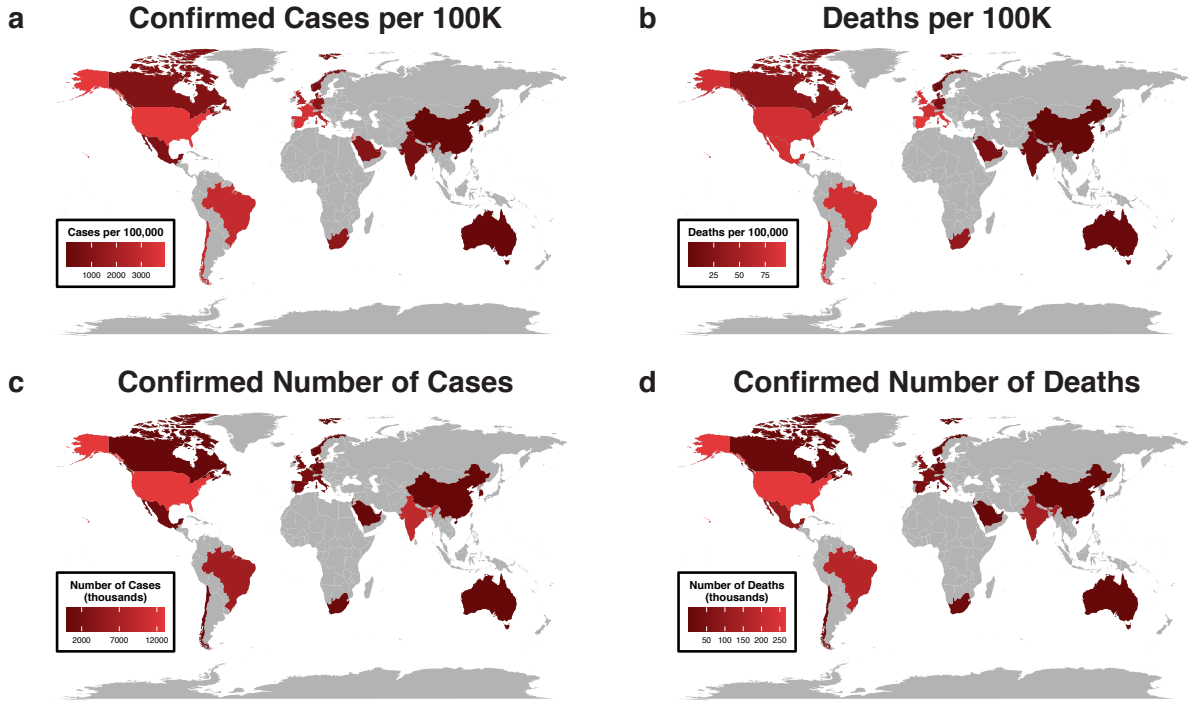
Spain	Lucid	Spanish	1000	1253
United Arab Emirates	Lucid	Arabic	500	734
United Kingdom	Prolific	English	1000	863
United States	Prolific	English	1000	886

Supplementary Table 2. Participant information for the self-report and voting task in each country. Table includes final sample size, mean age, participants' self-reported gender identity (as a woman, man, or another gender identity).

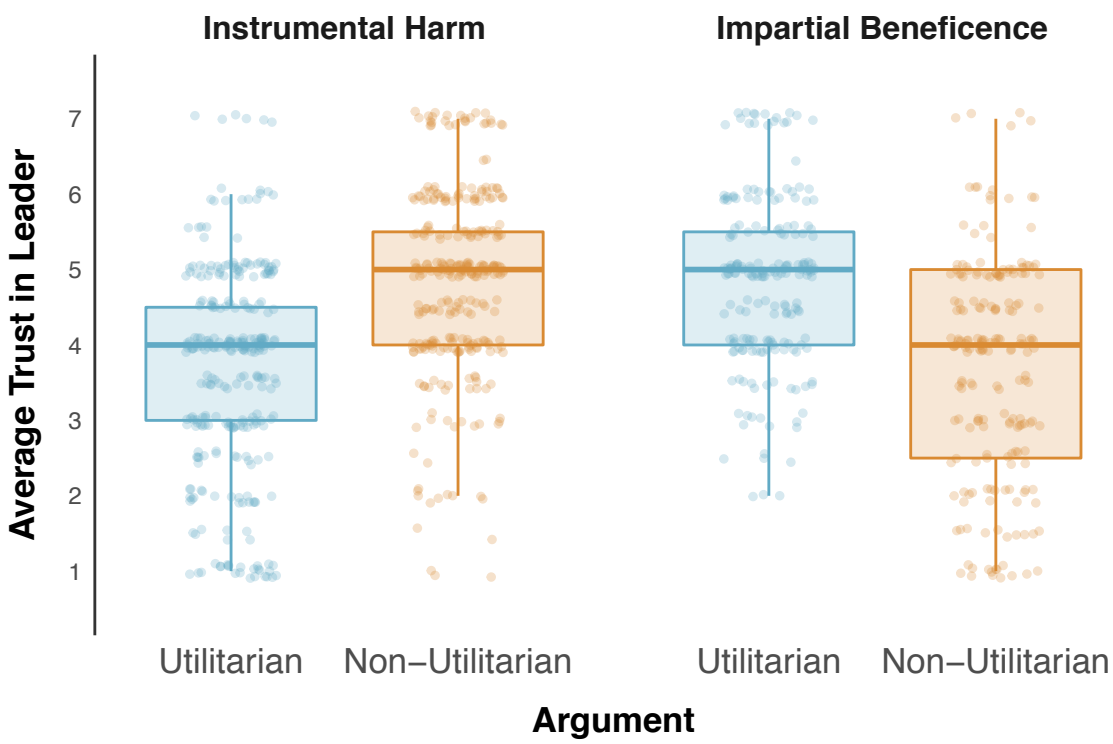
Country	Self-Report Task					Voting Task				
	N Total	Mean Age	Gender Identity			N Total	Mean Age	Gender Identity		
			Woman	Man	Other			Woman	Man	Other
Australia	790	47.59	445	341	4	574	47.92	321	250	3
Brasil	868	40.45	430	429	9	542	39.21	253	285	4
Canada	913	48.62	467	441	5	647	46.79	334	309	4
Chile	943	41.49	474	447	22	575	39.99	276	285	14
China	771	39.64	384	367	20	639	41.84	321	298	20
Denmark	883	46.95	459	421	3	648	47.34	328	319	1
France	878	48.74	474	402	2	638	47.77	322	315	1
Germany	990	50.83	466	518	6	741	51.75	342	396	3
India	843	37.92	424	417	2	777	38.83	361	413	3
Israel	711	43.00	353	353	5	328	40.20	162	165	1
Italy	932	48.56	450	476	6	360	44.73	175	184	1
Kingdom of Saudi Arabia	541	31.23	292	246	3	447	31.33	234	210	3
Mexico	777	38.76	404	364	9	600	38.35	289	304	7
Netherlands	828	47.48	444	377	7	750	47.59	401	342	7
Norway	1060	45.96	560	495	5	448	43.22	237	208	3
Singapore	536	35.63	243	289	4	443	35.30	209	229	5
South Africa	879	39.43	459	415	5	670	38.65	353	314	3
South Korea	453	45.34	238	215	0	409	46.20	194	214	1
Spain	846	47.53	388	453	5	542	45.53	237	303	2

United Arab Emirates	519	31.14	251	260	8	439	31.30	217	215	7
United Kingdom	809	45.93	423	382	4	685	46.05	357	325	3
United States	821	45.29	430	375	16	736	45.13	371	351	14

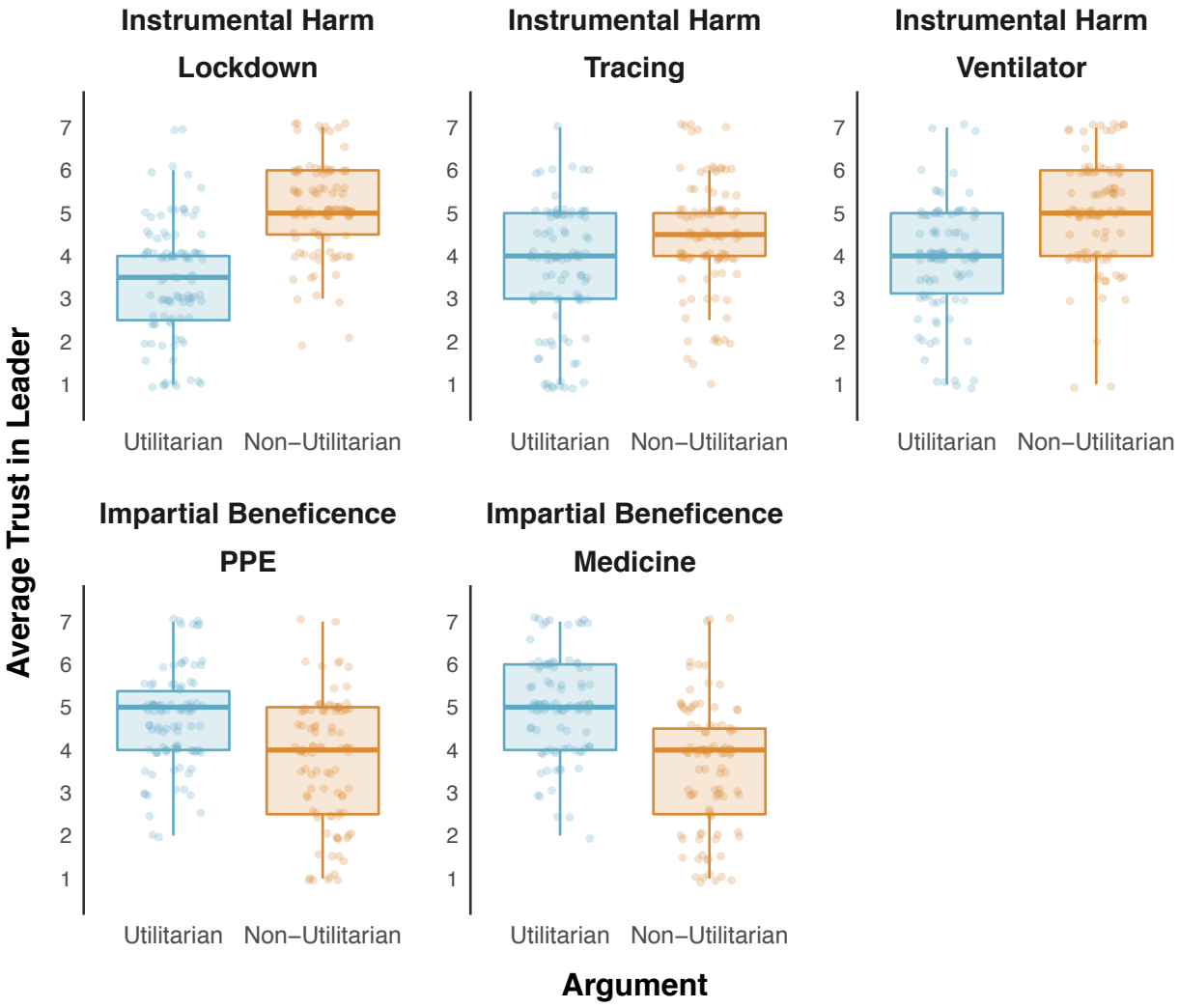
Supplementary Figures and Captions



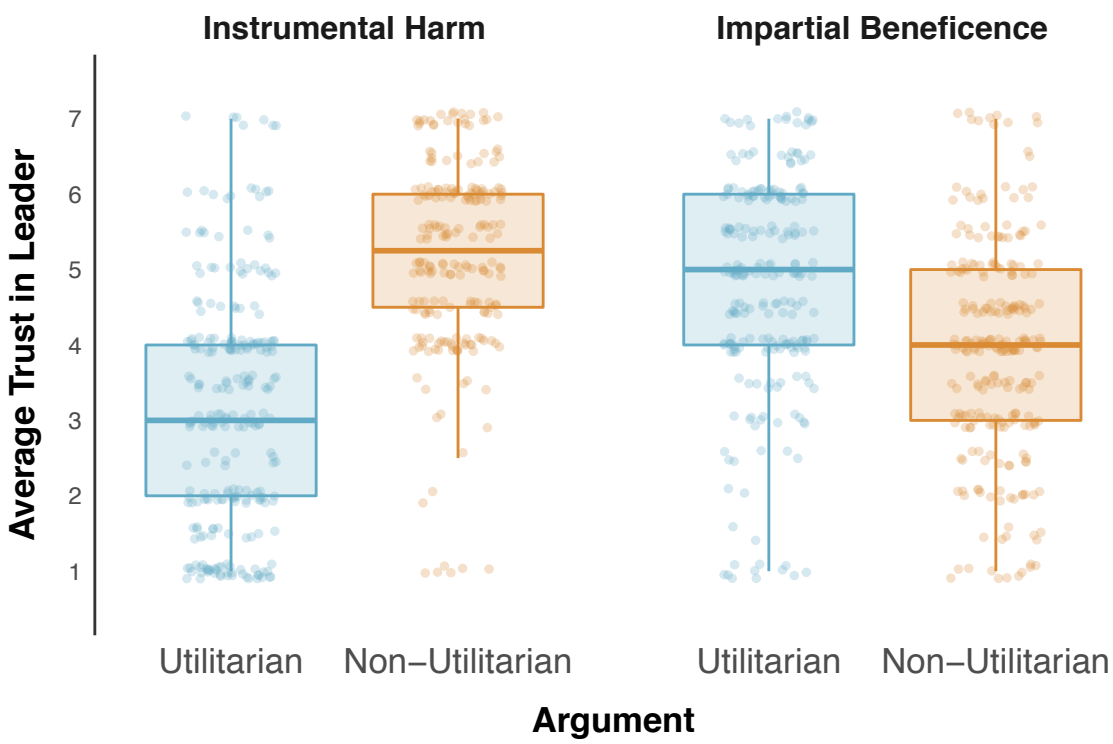
Supplementary Figure 1. Impact of the COVID-19 Pandemic on Countries of Recruitment at the Time of Data Collection. (A) Number of confirmed COVID-19 cases per 100 thousand people in each country of intended recruitment. (B) Number of COVID-19 deaths per 100 thousand people in each country of intended 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[3] (last update: November 26th, 2020). Population estimates for each country were taken from the United Nations' World Population Prospects (last update: July 1st, 2019).



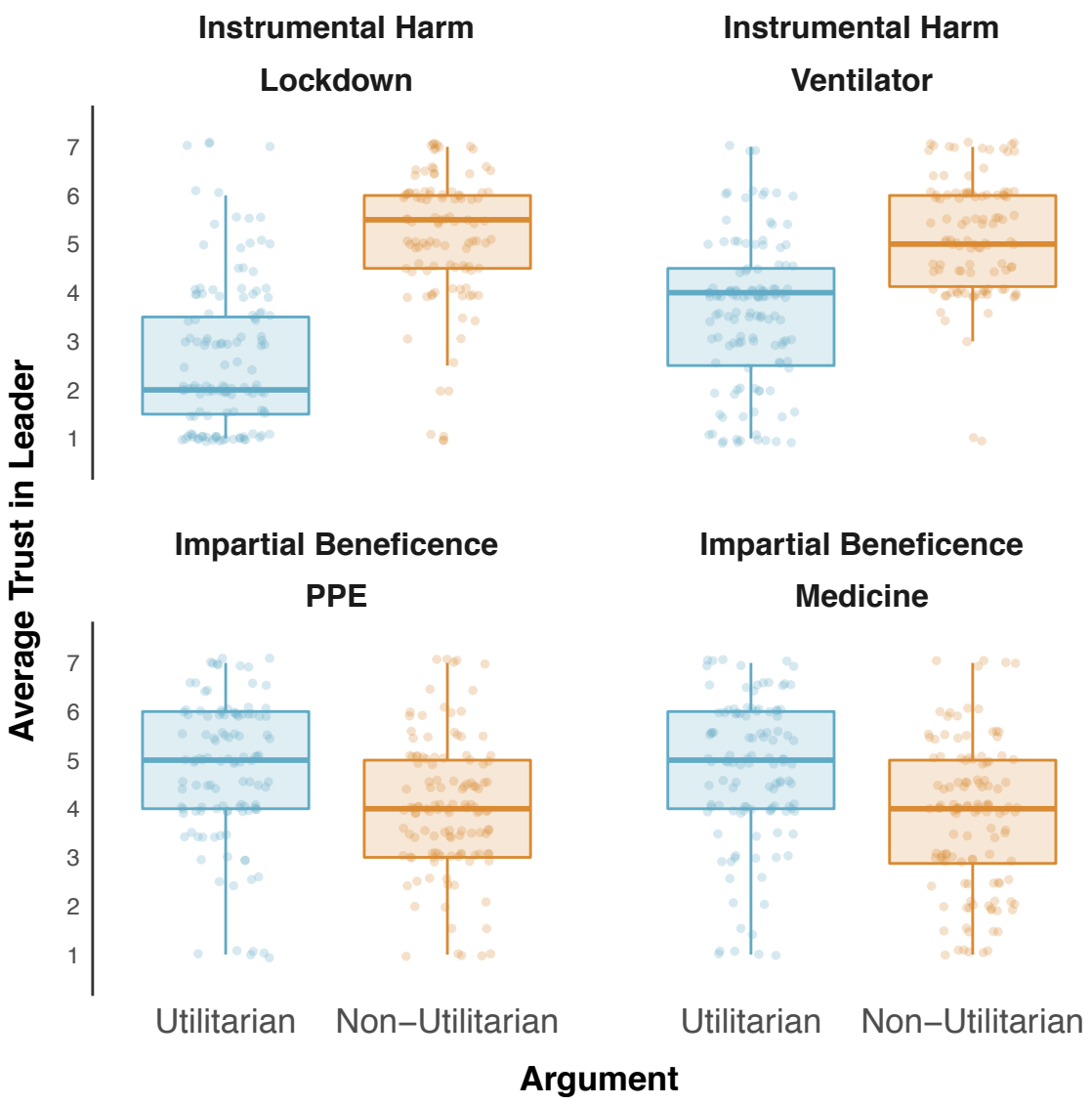
Supplementary Figure 2. Self Reported Trust in Pilot 1. Average self-reported trust in utilitarian vs non-utilitarian leaders in Pilot 1 (N = 98), separately for Instrumental Harm dilemmas (Lockdown, Tracing, and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders 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.



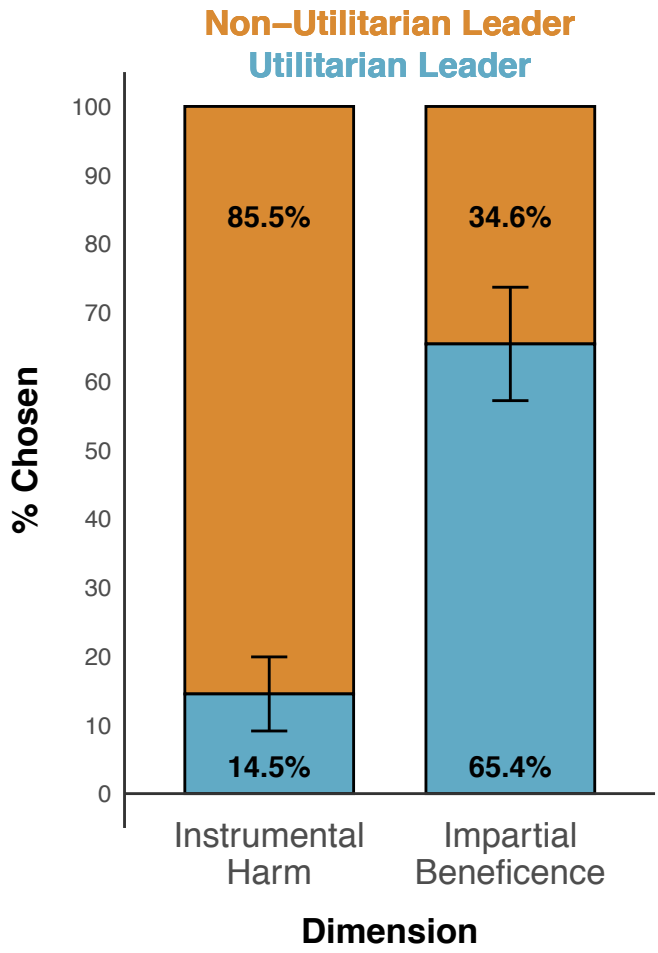
Supplementary Figure 3. Self Reported Trust by Dilemma in Pilot 1. Average self-reported trust in utilitarian vs. non-utilitarian leaders in Pilot 1, separately for each dilemma, including both Instrumental Harm dilemmas (Lockdown, Ventilators, and Tracing) and Impartial Beneficence dilemmas (Medicine 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.



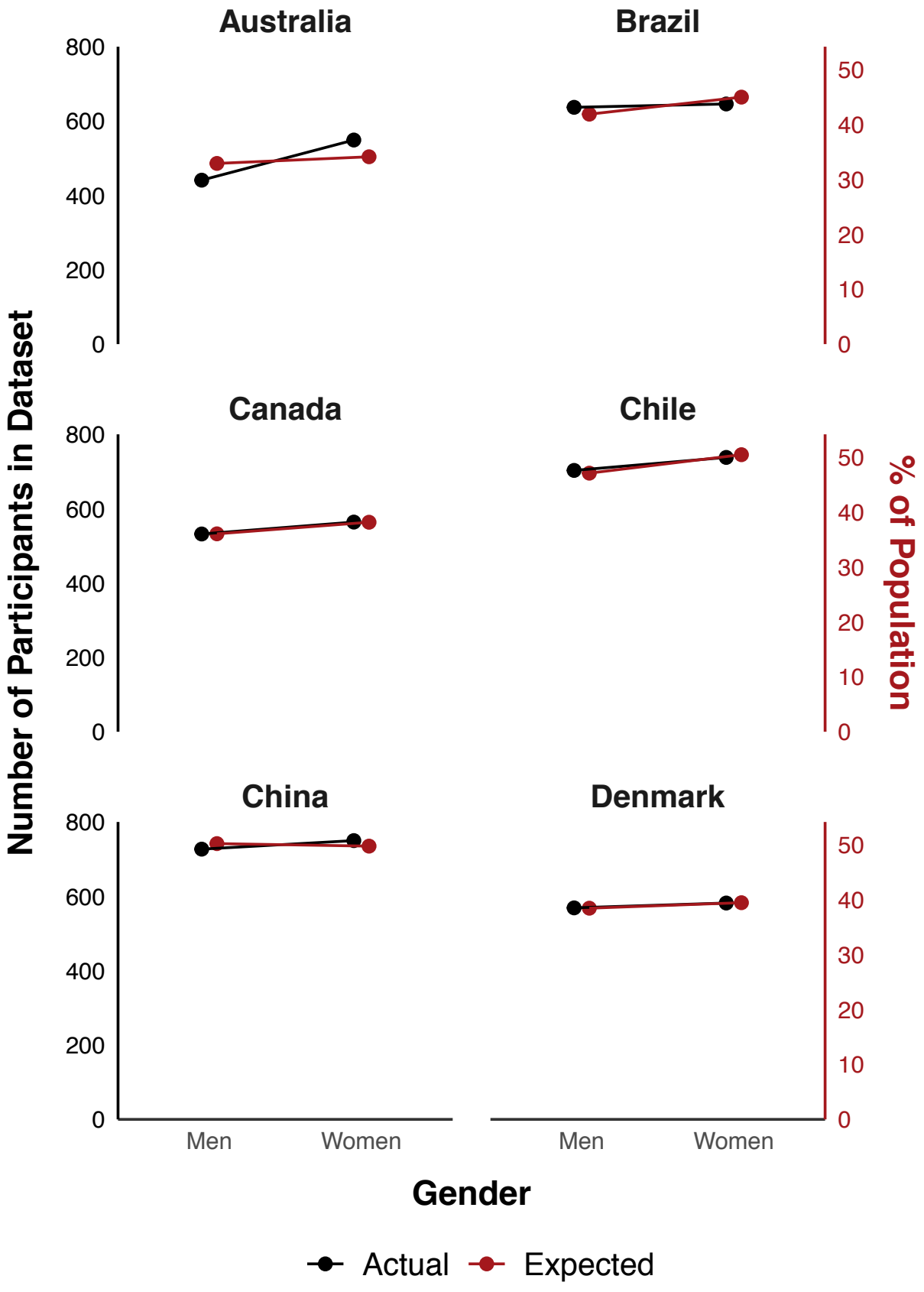
Supplementary Figure 4. ***Self Reported Trust in Pilot 2***. Average self-reported trust for the utilitarian vs. non-utilitarian leaders in Pilot 2 (N = 469), separately for Instrumental Harm dilemmas (Lockdown and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). Non-utilitarian leaders 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.

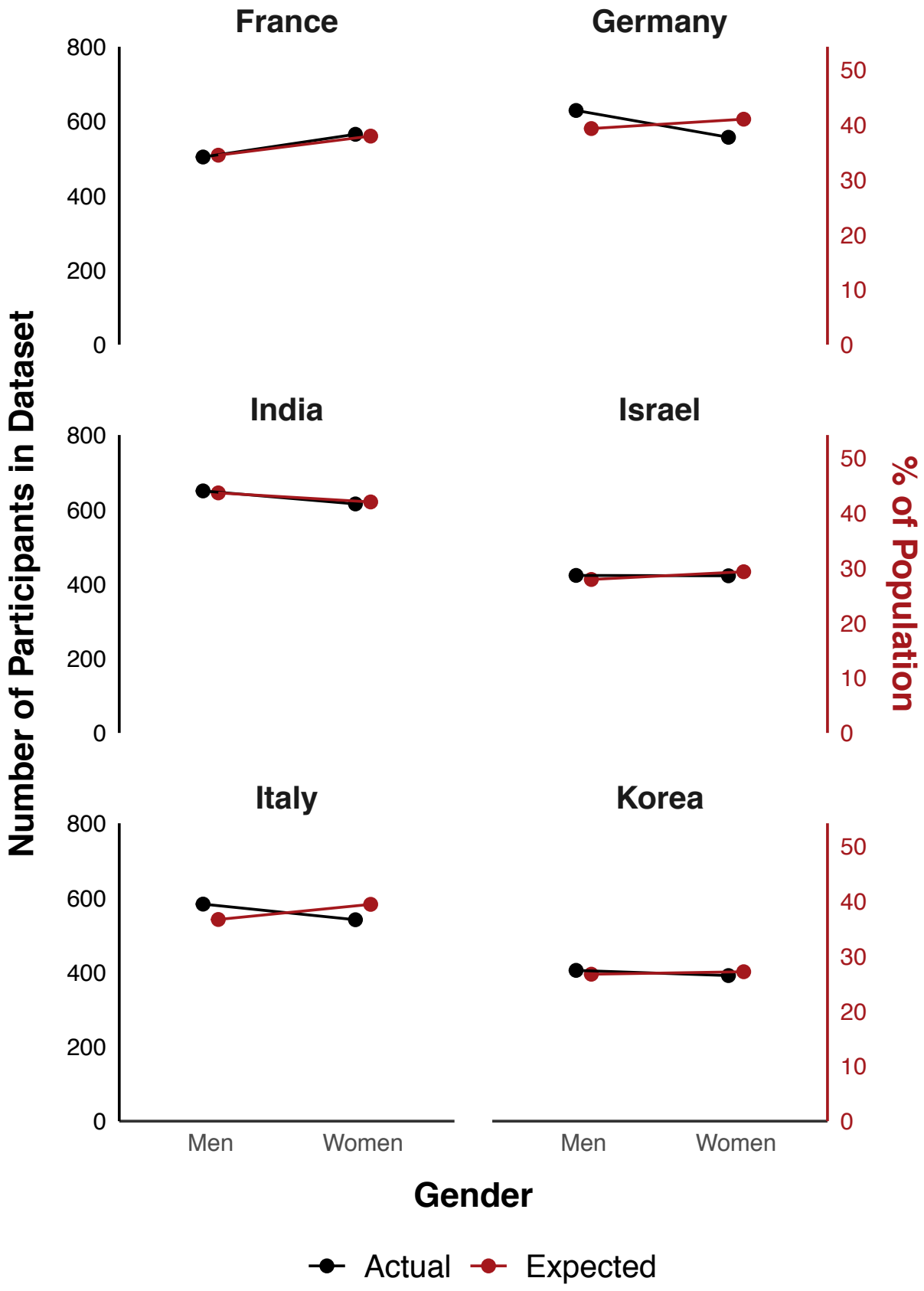


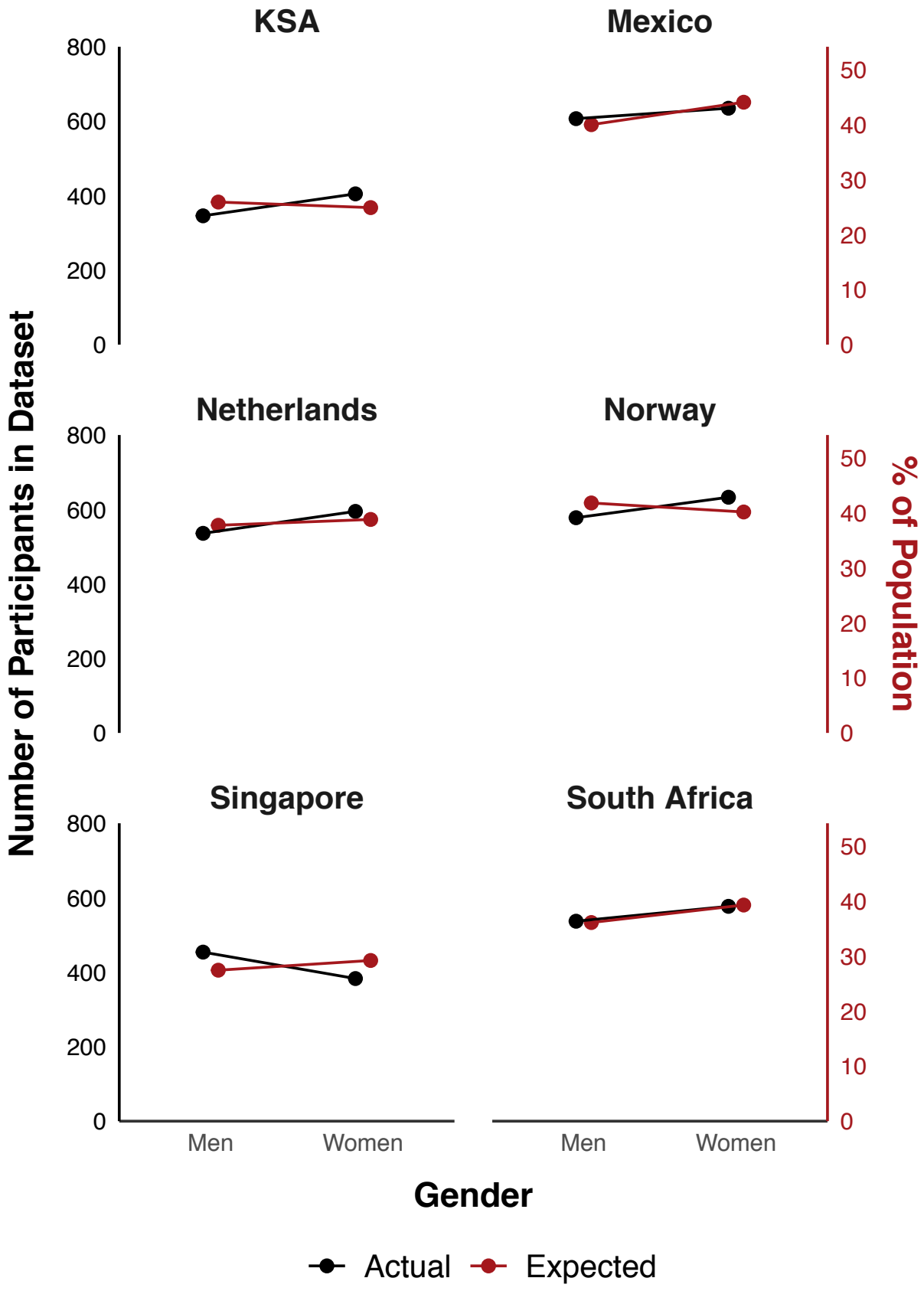
Supplementary Figure 5. Self Reported Trust by Dilemma in Pilot 2. Average self-reported trust in utilitarian vs non-utilitarian leaders in Pilot 2, separately for each dilemma, including both Instrumental Harm dilemmas (Lockdown and Ventilators) and Impartial Beneficence dilemmas (Medicine and PPE). 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.

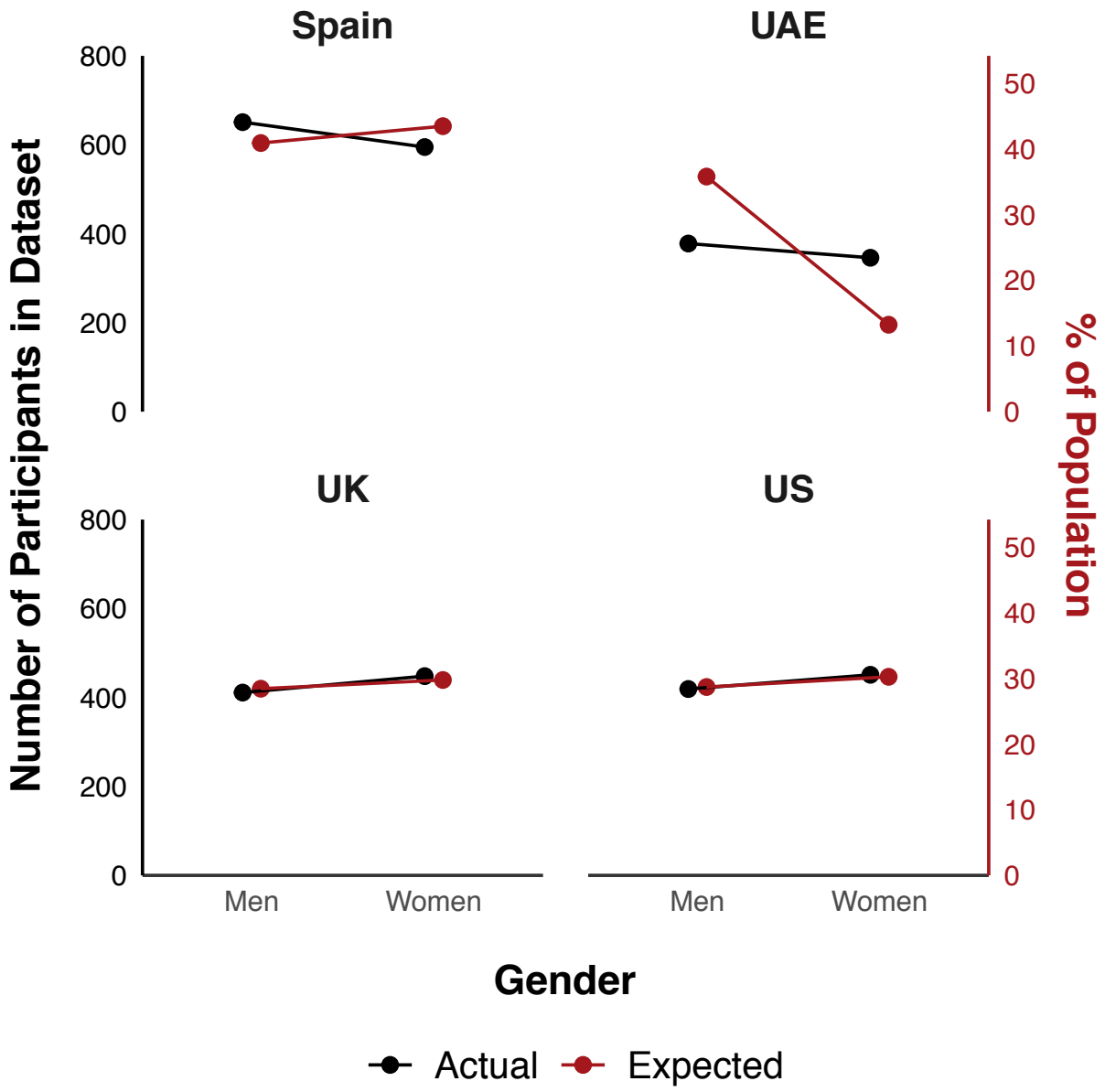


Supplementary Figure 6. ***Voting Choices in Pilot 2***. Model estimates of the percentage of participants who chose to entrust utilitarian vs. non-utilitarian leaders in the voting task in Pilot 2 (N = 452), separately for Instrumental Harm (Lockdown and Ventilators) and Impartial Beneficence dilemmas (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.

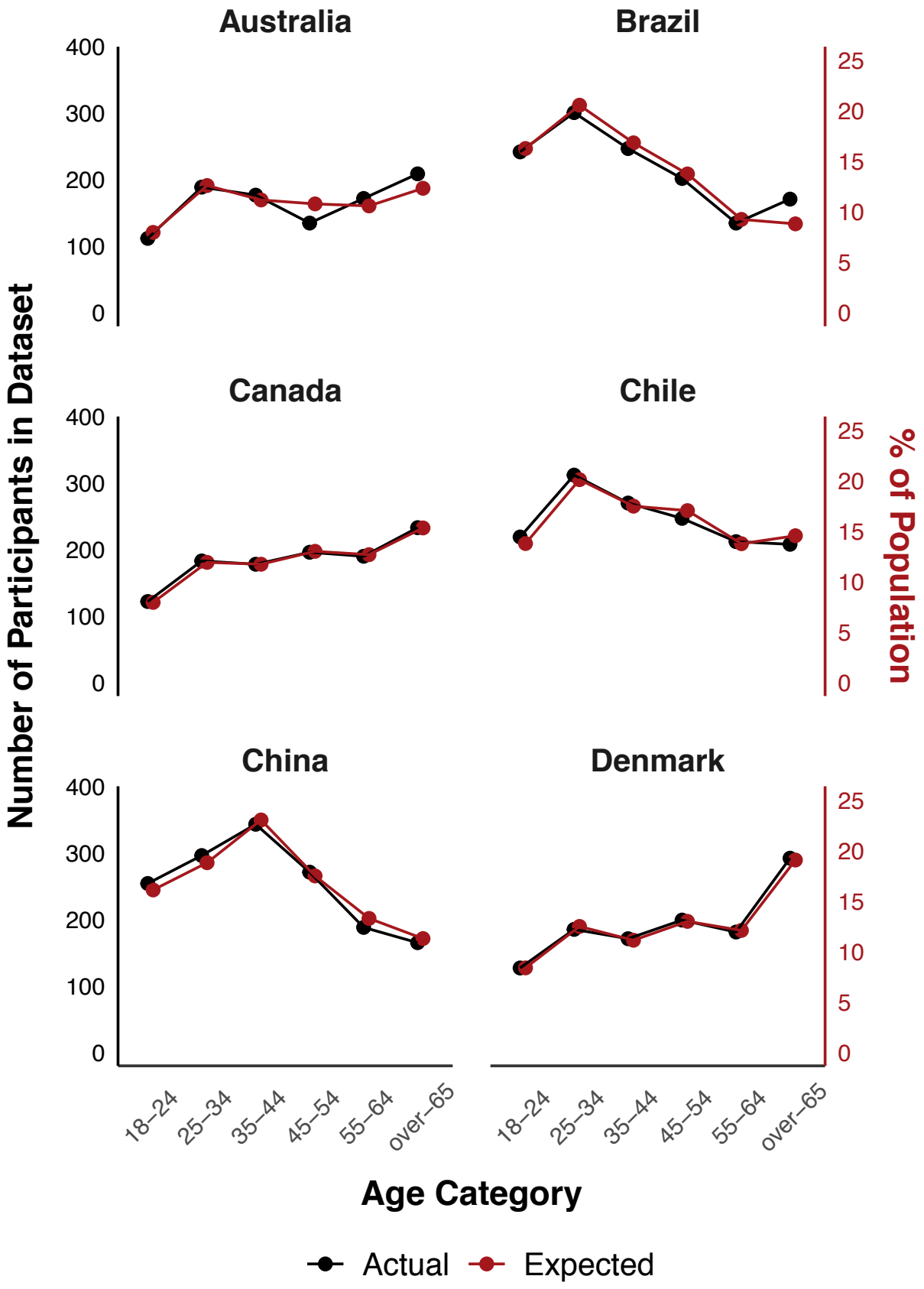

Australia
Brazil
Canada
Chile
China
Denmark
800
600
400
200
0
50
40
30
20
10
0
Number of Participants in Dataset
% of Population
Men
Women
Gender
Actual
Expected

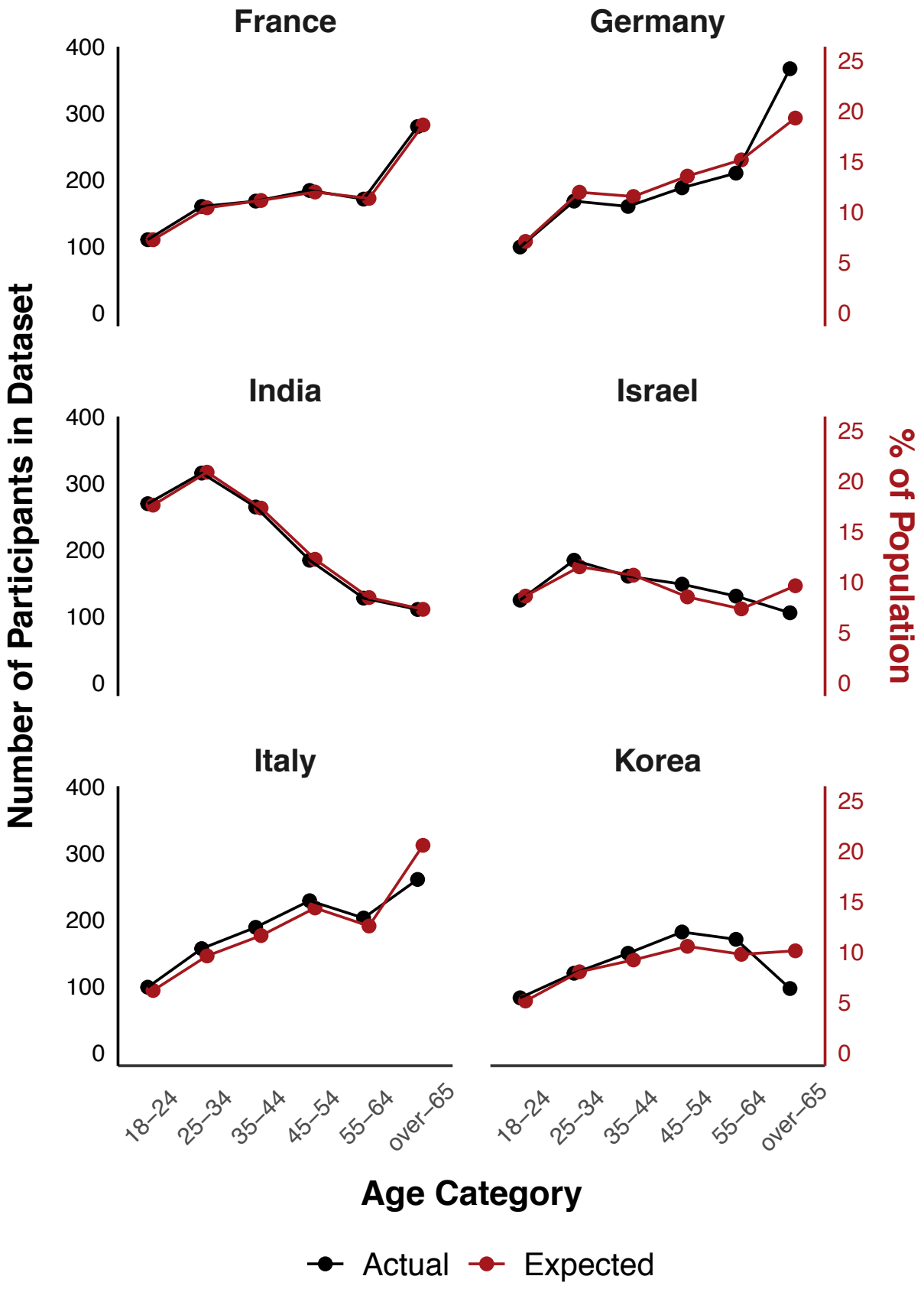

France
Germany
India
Israel
Italy
Korea
800
600
400
200
0
50
40
30
20
10
0
Number of Participants in Dataset
% of Population
Men
Women
Gender
Actual
Expected

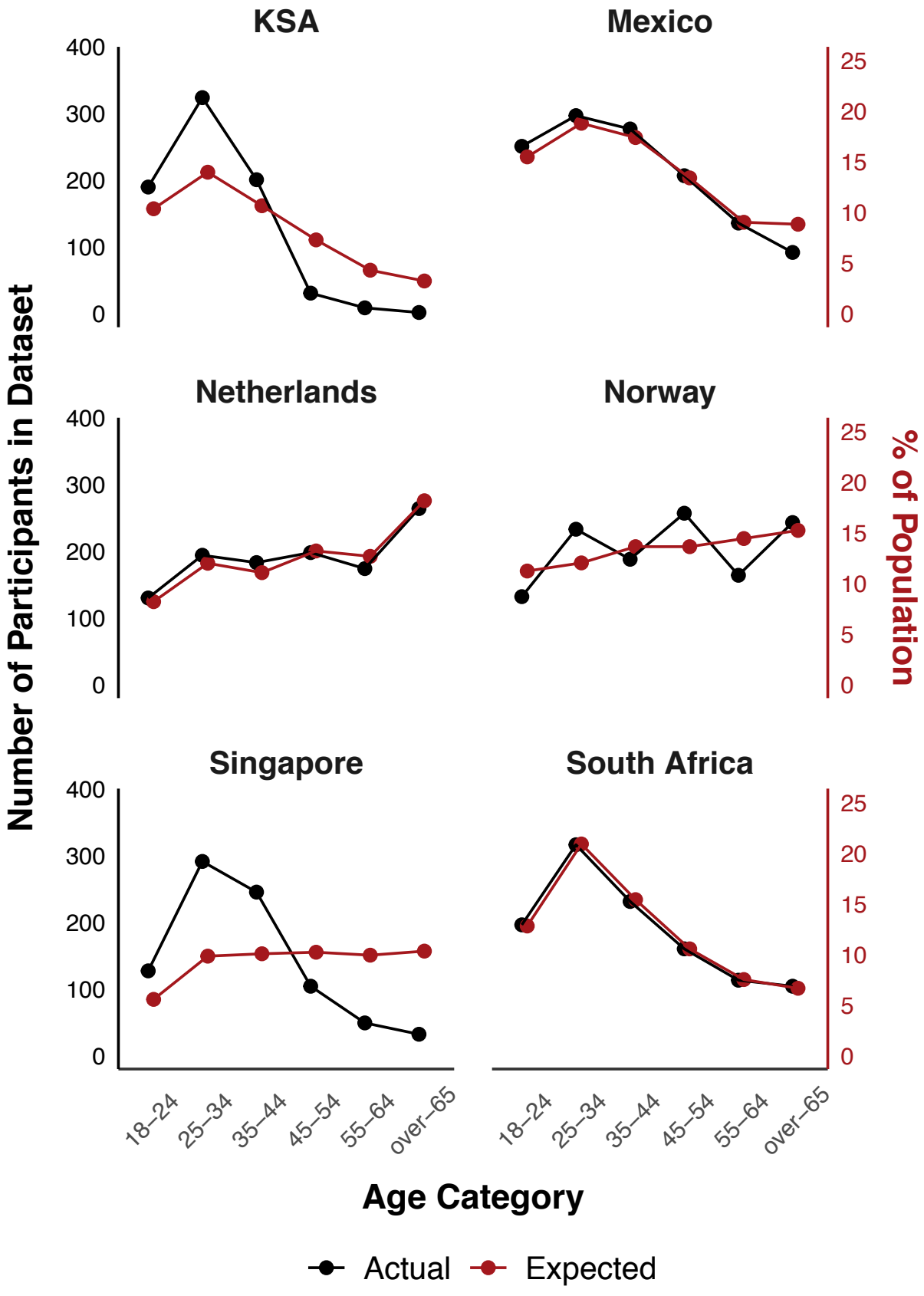

KSA
Mexico
Netherlands
Norway
Singapore
South Africa
800
600
400
200
0
50
40
30
20
10
0
Number of Participants in Dataset
% of Population
Men
Women
Gender
Actual
Expected

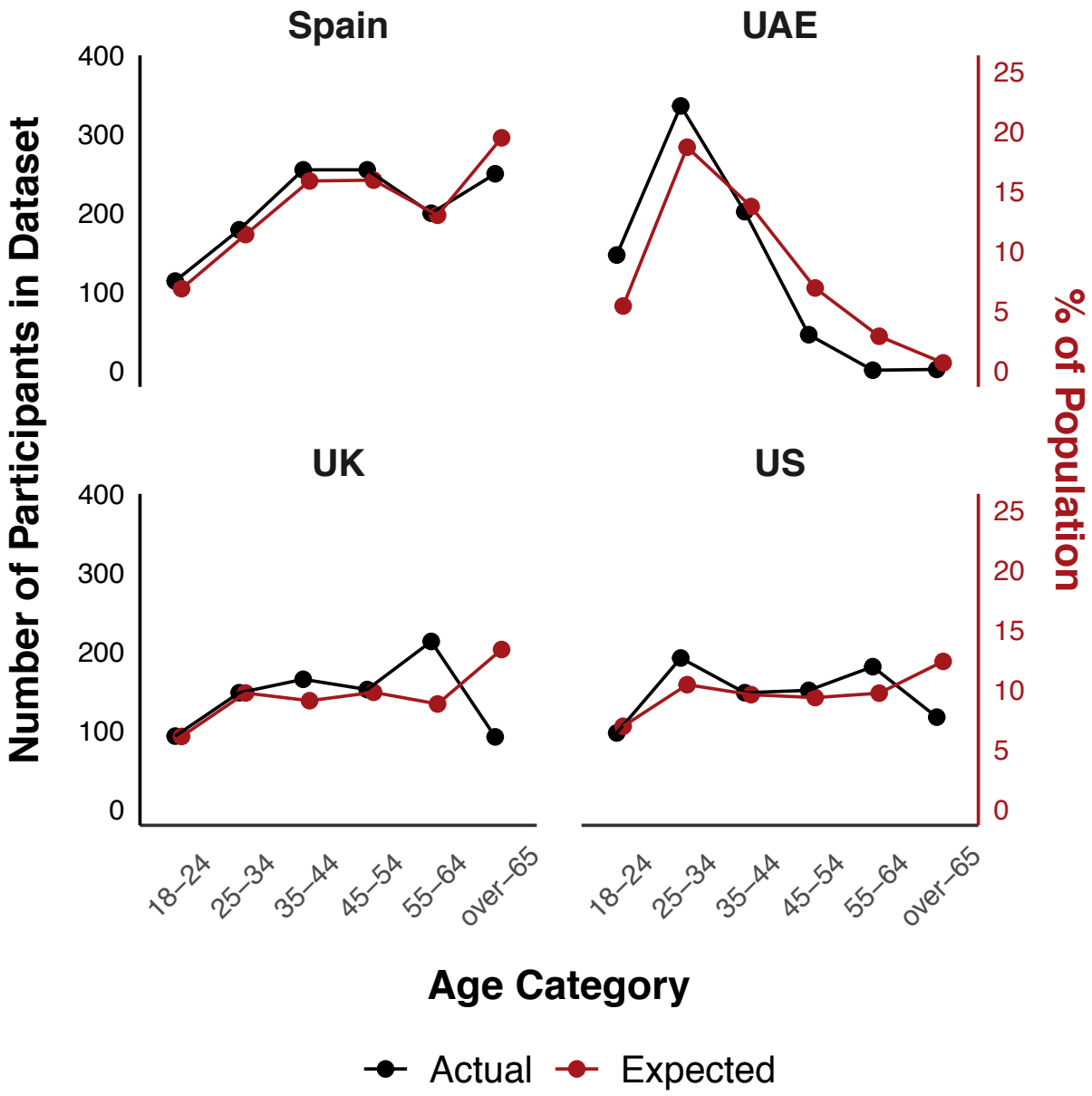


Supplementary Figure 7. Actual vs. Expected Number of Men and Women in Each Country. Black lines indicate the final number of women and men in our dataset (across both self-report and voting tasks), while red ones indicate the expected number of women and men based on each country's population characteristics.


Australia
Brazil
Canada
Chile
China
Denmark
Number of Participants in Dataset
% of Population
0
100
200
300
400
0
5
10
15
20
25
18–24
25–34
35–44
45–54
55–64
over–65
Age Category
Actual
Expected


France
Germany
India
Israel
Italy
Korea
Number of Participants in Dataset
% of Population
0
100
200
300
400
0
5
10
15
20
25
18–24
25–34
35–44
45–54
55–64
over–65
Age Category
Actual
Expected


KSA
Mexico
Netherlands
Norway
Singapore
South Africa
Number of Participants in Dataset
% of Population
400
300
200
100
0
25
20
15
10
5
0
18–24
25–34
35–44
45–54
55–64
over–65
Age Category
Actual
Expected



Supplementary Figure 8*. *Actual vs. Expected Number of Participants for Age Categories in Each Country. Black lines indicate the final number of participants for each age category in our dataset (across both self-report and voting tasks), while red ones indicate the expected number in each age category based on each country's population characteristics.

References

1. Barr, D. J. Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* **4**, (2013).
2. Gomila, R. Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis. *J. Exp. Psychol. Gen.* **150**, 700–709 (2021).
3. Wang, Y. A., Sparks, J., Gonzales, J. E., Hess, Y. D. & Ledgerwood, A. Using independent covariates in experimental designs: Quantifying the trade-off between power boost and Type I error inflation. *J. Exp. Soc. Psychol.* **72**, 118–124 (2017).
4. Wysocki, A., Lawson, K. M. & Rhemtulla, M. Statistical Control Requires Causal Justification. (2020) doi:10.31234/osf.io/j9vw4.
5. Urminsky, O., Hansen, C. & Chernozhukov, V. Using Double-Lasso Regression for Principled Variable Selection. SSRN Electron. J. (2016) doi:10.2139/ssrn.2733374.