Altruism in governance: Insight from randomized training

Sultan Mehmood (New Economic School), Shaheen Naseer (Lahore School of Economics) and Daniel L. Chen (Toulouse School of Economics)

Global Priorities Institute | July 2022

GPI Working Paper No. 7-2022

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

Altruism in Governance: Insights from Randomized Training

BY SULTAN MEHMOOD, SHAHEEN NASEER AND DANIEL L. CHEN[1]

Randomizing different schools of thought in training altruism finds that training junior deputy ministers in the utility of empathy renders at least a 0.4 standard deviation increase in altruism. Treated ministers increased their perspective-taking: blood donations doubled, but only when blood banks requested their exact blood type. Perspective-taking in strategic dilemmas improved. Field measures such as orphanage visits and volunteering in impoverished schools also increased, as did their test scores in teamwork assessments in policy scenarios. Overall, our results underscore that the utility of empathy can be a parsimonious foundation for the formation of prosociality, even impacting the behavior of adults in the field.

JEL Codes: J01, J24; D64; D73; M50

[1]Sultan Mehmood: New Economic School (E-mail: smehmood@nes.ru), Shaheen Naseer: University of Oxford (Email: shaheen.naseer@oxfordmartin.ox.ac.uk), Daniel Chen: Toulouse School of Economics (E-mail: daniel.chen@iast.fr). AEA RCT Registry number ID is: AEARCTR-0006655. We would like to thank Peter Singer, John List, Ray Fisman, Kevin Tobias, Moshe Hoffman, Jamil Zaki, Rosemarie Nagel, Martha Nussbaum, and seminar participants at Institute for Advanced Study in Toulouse, Paris School of Economics, Aix-Marseille University, American Political Science Association, NBER Fall Development Meeting 2021 and World Bank for their helpful suggestions and comments. We would also like to thank Chris Wickens and Benson Njogu for their invaluable assistance in coding the experiments on oTree. Shicheng Xia, Sajwaar Khalid, Shilpa Rao, and Sameen Tariq provided excellent research assistance. Daniel Chen acknowledges IAST funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EUR-0010. This research has also benefited from financial support of the research foundation TSE-Partnership. An IRB is obtained with IRB Number RERC-022021-01.

We can never survey our own sentiments and motives, we can never form any judgment concerning them; unless we remove ourselves, as it were, from our own natural station, and endeavour to view them as at a certain distance from us. But we can do this in no other way than by endeavouring to view them with the eyes of other people, or as other people are likely to view them. —Adam Smith, *The Theory of Moral Sentiments* (1759)

Prosociality—behavior that benefits others or society as a whole—is critical in contract enforcement, management of commons, public goods provision, establishing effective rule of law, efficient governance in societies and for labor market success (Knack and Keefer, 1997; La Porta et al., 1997; Fehr and Gächter, 2002; Ostrom et al., 2002; Henrich et al., 2004; Guiso et al., 2009; Bloom and Van Reenen, 2011; Cooper and Kagel, 2016; Burks et al., 2016; Robalino and Robson, 2016; Deming, 2017; Kosse and Tincani, 2020). This raises an important question: How can prosociality be cultivated? Beyond laboratory studies showing the short-term malleability of prosocial behavior, there have been few field experiments that look at how to train prosociality effectively, especially in adults. A pioneering experiment found improvements in prosociality after an early childhood intervention (Heckman et al., 2013), while recent experiments build on this study and found improvements in prosocial behavior through mentoring elementary school children for one year (Falk et al., 2020) and from a yearlong, three-hour-per-week curriculum designed to build social cohesion in schools (Alan et al., 2021). We explore a parsimonious and scalable way to train prosociality among adults. We pre-register a randomized control trial of different schools of thought from economics and from psychology on cultivating prosociality and test whether emphasizing the utility of empathy as opposed to emphasizing the malleability of the self helps cultivate prosociality. To build prosociality, we leverage recent economic insights on the increasing importance of soft skills[2] – empathy in particular (Deming, 2017). Existing literature supports the connection between empathy and prosocial behavior, as well as between perspective-taking and prosocial behavior (Eisenberg and Miller, 1987; Eisenberg et al.,1991). Perspective-taking or "putting oneself in another's shoes" (Premack and Woodruff, 1978) is often called the "Theory of Mind" by psychologists and the "Degree of Strategic Reasoning" by economists. Soft skills have been formally modeled to reduce coordination costs so that teams, organizations, and society can work together

[2] Soft skills, also called non-cognitive skills, are simply the residual that is not predicted by IQ or achievement tests (Deming, 2017). Soft-skills include skills like emotional intelligence, collaboration, teamwork, and empathy.

more effectively, but there are two challenges: 1) measuring soft skills (such as teamwork and coordination) and 2) identifying causal effects (Deming and Weidmann, 2021). Our paper seeks to make progress on these challenges.

We conducted a randomized evaluation with junior deputy ministers. To cultivate prosociality, we randomized junior ministers into four training workshops. The first training workshop emphasizes the utility of empathy, with a focus on how empathy influences overall organizational and individual performance. The lecture focused on narratives on how empathy was a skill to get ahead in ministers' careers and presented quantitative evidence from the private sector on how empathic behavior is beneficial. The second group of ministers was randomized into the malleability of empathy workshop, where the emphasis was on empathy being mutable and subject to growth. This message too, was delivered by presenting narratives, but these narratives showcased individuals growing in empathy. It also reported quantitative evidence from the private sector that empathy is malleable. In the third training workshop, we combined the key messages from both training workshops, emphasizing both the benefits and malleability of empathy. We evaluated these three training programs against a placebo training in macroeconomics, which was unrelated to empathy. The macroeconomics workshop presented basic concepts in macroeconomics, including facts about the macroeconomics of Pakistan and a generic discussion of GDP, GNP, inflation, and unemployment.

The experiment involves five stages. Stage I was a 15-minute recorded lecture, followed by a short writing exercise that covered the main lessons learned in the lecture. Two weeks later, Stage II took place. Stage II consisted of a 2-hour live Zoom session in which the junior ministers first participated in a 10-minute structured discussion about their previous assignment[3] and then listened to a 50-minute lecture on the importance of emotional intelligence. Then Stage II ended with participants engaging in a 1-hour interactive activity session that consisted of playing 12 behavioral games. Stage III began by measuring empathy in the field two months following the interventions through the solicitation of blood donations. Stage IV involved a book assignment where the junior ministers were cross-randomized to either receive empathy or a placebo book. The ministers then wrote two 1500-word essays on the main lessons of their assigned book and its application to their future career. Finally, four and six months post-lecture respectively, Stage V commenced. In Stage V, we

[3] The structured discussions were carried out in breakout rooms prior to the main lecture so only those assigned to their treatment condition, U, M, UM, or placebo would discuss the State I material with each other.

investigated the impact of the treatments on field visits (four months post) and regular assessments (six months post).

To assess the impact of the workshops on prosocial behavior, we developed a set of indicators: (i) non-incentivized lab-in-field experiments to assess prosocial behaviors such as altruism, coordination, cooperation, and, perspective taking; (ii) responses to donate blood by the junior ministers; (iii) "field trip" records of orphanage and school visits; and (iv) official assessments such as teamwork, quantitative research methods, and soft skills.

We measure altruism in the laboratory (donations given to each other and to charities). First, we measure perspective-taking in strategic dilemmas: cooperation and coordination. Past studies have documented that high performance in these strategic dilemmas is associated with neural activity in the medial prefrontal cortex which is associated with successful mentalizing (Coricelli and Nagel, 2009). The values encoded in a region of the prefrontal cortex are causally related to economic choices (Ballesta et al., 2020). We also observe honesty in the die-rolling or "lying game" (Abeler, et al., 2019; Gneezy, et al., 2018; Fischbacher, et al. 2013). Second, volunteers from a prominent blood bank made one of two types of blood donation requests to the ministers. One type specifically asked for the minister's matching blood type, while the other was a general request for blood donations. The ministers were cross-randomized to receive one of the two requests. Our third measure of prosocial behavior attempts to capture the impact of our treatments on actual behavior that is measured after four months of the intervention. Fourth, we examine the impact of our treatment on regular assessments on teamwork, research methods and soft skills to see if the laboratory measures translate to measures in the field.

Our findings show that junior ministers assigned to the utilitarian training workshop exhibited higher levels of altruism, improved perspective-taking, and increased blood donations compared to the control group. We find that blood donations increased, especially when the deputy ministers were told that their exact blood type was in need. The utilitarian group also demonstrated a rise in orphanage visits and volunteering to teach at underprivileged schools. Furthermore, the group received higher scores in their regular evaluations of soft skills and teamwork, while their assessments on quantitative research methods assessments remained unchanged. Our results suggest that the utilitarian training had a positive impact on the junior ministers in areas beyond those measured in the lab-in-the-field setting.

Conversely, we did not observe any significant changes in the outcomes measured for the malleability of empathy workshop or the joint training, which combined the utilitarian and malleability of empathy. We interpret this null result of the malleability and joint treatment in light of the theoretical self-image models of Benabou and Tirole (2004, 2006, 2011). While the predictions of the model are theoretically ambiguous because the treatments can affect different parameters of the model, the dominant channel through which the utilitarian training appears to have an effect is the extrinsic value of acting prosocially. In this framework, empathetic behavior also informs our identity as a prosocial person. The malleability of one's prosociality means that our behavior is less informative about our identity. Formally, utilitarian training increases the private benefits of empathy while malleability training may have its dominant effect through reducing the updating of self-perceptions upon taking empathetic actions. Consistent with this, we find deputy ministers treated with the malleability of the self decreased their ratings on the importance of prosociality.

The paper contributes to several strands of literature in economics, psychology, and philosophy. First, to the best of our knowledge, we are the first to show that altruism training can impact behavior in adults. As such, our study is related to the formation of prosociality (Kautz et al., 2014; Kosse et al., 2020; Lindauer et al., 2020). A few randomized control trials that also find effects of training interventions (Heckman et al., 2013; Falk et al., 2020; Alan et al., 2021; Cappelen et al., 2020). Our results suggest that a utilitarian empathy workshop could provide an economical foundation for the formation of prosociality in adults. This would be consistent with evolutionary theories on the formation of prosociality that suggest that prosociality is plastic and mutable (Francois et al., 2018). Second, we contribute to the literature on soft skills, which labor economists recognize as explaining large puzzles in the labor market over the last half-century (Autor et al., 2015; Deming, 2017). Soft skills are also likely a key ingredient in the personnel economics of the state (Finan et al., (2017). A recent literature review highlighted three important channels for improving public service in developing countries—selection, incentives, and monitoring (Finan et al., 2017)—but there was no attention paid to soft skills nor how these "technologies" of production can be enhanced after the recruitment of public officials. To be clear, changing any of these factors – selection, incentives, monitoring, and even soft skills can theoretically decrease social welfare (Ashraf et al., 2020); however, we find evidence consistent with an increase in social welfare. For instance, teaching people about the private benefits of empathy in our utilitarian training group led to increases in blood

donations in a context and time when "blood banks were practically empty" (Shaukat Khanam Hospital, 2021).

Third, we show that training the utilitarian value of empathy can impact field behavior. We build on recent online survey experiments estimating the impact of training ideas associated with rational appeal can impact charitable donations (Lindauer et al., 2020). We complement this important study as our work teaches the utility of empathy in the field, with deputy ministers, and traces their impact on both prosocial behavior in the field (donations of blood and time) and performance in ministers policy exams. As such, our study complements recent theoretical developments in modeling the motivations of high-stakes decision-makers such as public servants and politicians, where self-image and prosocial behavior may be an important driver of effective service delivery (Besley and Ghatak, 2018; Barfort et al., 2019; Gulzar and Khan, 2021; Ashraf et al., 2020). We also map competing schools of thought (utilitarian vs malleability of empathy) on cultivating prosociality into these formal models and test them empirically.

This paper is organized as follows. Section I provides background information and the set-up of the experiment. Section II describes the data and empirical strategy. Section III presents the results from the lab and the field. Section IV concludes.

I. Background, Theoretical Framework, and the Study Design

A. *Background*

The deputy ministers hold important positions in district administration, federal and local ministries, central government secretariats, and public enterprises. This system is like those in India and other common law countries with colonial supplantation of institutions (see e.g. Iyer and Mani, 2012). The government considers these policymakers as "key wheels on which the entire engine of the state runs" so these are high-stakes decision-makers. These deputy ministers are selected through competitive examinations. The deputy ministers participate in regular training programs. The training involves participating in workshops on various subjects such as public sector management, politics, history, economics, and professional etiquette.

B. Theoretical Framework

The framework of self-image models from Benabou and Tirole, (2006) puts the utilitarian and malleability treatments in contrast. Denote an agent's intrinsic motivation and extrinsic motivation for prosocial or empathetic behaviours by v_a and v_y. The agent chooses a prosocial activity participation level a from some choice set $A \subset R$, which thus induces a utility cost $C(a)$, while it yields an extrinsic (possibly monetary) payoff y. Additionally, the agent's participation level would also signal his/her prosocial identity to others in the society, from which the agent derives a reputational payoff $R(a, y) = \mu_a E(v_a|a, y)$ with $\mu_a > 0$ capturing to what extent the agent would like to demonstrate a prosocial self-image identity. The agent thus faces the utility maximization problem:

$$\{(v_a + v_y y)a - C(a) + R(a, y)\}$$

The first-order condition for an agent's choice of a is:

$$C'(a) = v_a + v_y y + \mu_a \frac{\partial E(v_a|a, y)}{\partial a}$$

Here we can adopt a specification of the model that builds on the familiar normal-learning setup. Let actions vary continuously over $A = R$, with the cost function being $C(a) = ka^2$ where $k > 0$. Also assume that everyone has the same image concern $\underline{\mu}_a$. The agent's (v_a, v_y) are drawn from:

$$\left(v_a \; v_y \right) \sim N\left(\underline{v}_a \; \underline{v}_y \;\; , \left[\sigma_a^2 \; \sigma_{ay} \; \sigma_{ay} \; \sigma_y^2 \right] \right)$$

Standard results for normal random variables then yield:

$$E(v_a|a, y) = \underline{v}_a + \rho(y) \cdot \left(ka - \underline{v}_a - \underline{v}_y y - \underline{\mu}_a \frac{\partial E(v_a|a, y)}{\partial a} \right),$$

where

$$\rho(y) = \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2}$$

Intuitively, the posterior assessment of an agent's intrinsic motivation, $E(v_a|a, y)$, is a weighted average of the prior $\underline{v}_a$ and of the marginal cost of his/her observed contribution, net of the average extrinsic and reputational incentives to contribute at that level.

Consider the benchmark case of no correlation ($\sigma_{ay} = 0$) such that $\rho(y) = \frac{1}{1+y^2\sigma_y^2/\sigma_a^2}$. Here we can consider $\theta \equiv \sigma_y/\sigma_a$ as the noise-to-signal ratio for the observers to determine the

agent's type. There is a unique equilibrium, in which an agent with preferences (v_a, v_y) contributes at the level:

$$a = \frac{v_a + v_y y}{k} + \underline{\mu}_a \rho(y) = \frac{v_a + v_y y}{k} + \frac{\underline{\mu}_a}{1 + \theta^2 y^2}$$

The Utility of Empathy Treatment (U training), which emphasizes the extrinsic payoff to prosocial behaviours, can be considered as an amplifier for y. As long as the agent's extrinsic motivation is above a certain threshold, that is:

$$v_y > \frac{2k\underline{\mu}_a \theta^2 y}{(1 + \theta^2 y^2)}$$

then the agent's prosocial activity participation level a would increase with y, since:

$$a'(y) = \frac{v_y}{k} - \frac{2\underline{\mu}_a \theta^2 y}{(1 + \theta^2 y^2)} > 0$$

The Malleability of Empathy Treatment (M training), with its emphasis on how the intrinsic motivation for prosociality can be amended and that it is not fixed, however, casts doubt on the effectiveness of signaling one's image intrinsically prosocial via studying someone's public image or prosocial actions such as blood donation and orphanage visit. The M training can be thus considered as contributing to the perceived noise-to-signal ratio θ for others to assess the agent's type. The fact that identity is malleable simply through a mindset intervention like ours can increase the noise-to-signal ratio. Thereby the agent's participation in prosocial activity would actually be decreasing with higher θ as:

$$a'(\theta) = -\frac{2\underline{\mu}_a y \theta^2}{(1 + y^2 \theta^2)} < 0$$

As a result, the joint treatment may have qualitatively different effects from the utilitarian treatment and also different effects than would be suggested by a reduced form analysis of the two treatments considered separately.

C. *Study Design*

Sample and Randomization.— The study took place with all 213 public officers. To the best of our knowledge, none of the participants had taken part in any prior randomized evaluation to the best of our knowledge. Our pre-registration was brief following recent suggestions in (Banerjee et al., 2020) for moderation in pre-analysis plans, so we registered the study design and the broad classes of outcomes: social preferences, bureaucratic performance, and thought leadership. In this paper, we focus on the first two classes of outcomes. The 213 deputy ministers were randomly assigned to one of the four treatment arms using a random number generator: (i) utilitarian treatment (53 participants); (ii) malleability treatment (54 participants); (iii) joint utilitarian and malleability treatment (53 participants) and (iv) placebo (53 participants).[4]

The Rollout.—The four treatments were delivered via a non-shareable and non-downloadable link containing four different training lectures.[5] The content for the training could only be accessed by entering the unique email address of the participant. We made sure that the training link was non-downloadable and could only be opened by the randomly assigned participant according to their treatment status.[6] The junior ministers were randomized into four training workshops. The first training workshop emphasized the value of empathy, that being empathetic is in the best interest of deputy ministers (n=53). The second training workshop focused on the concept of empathy's malleability, emphasizing that growth in empathy is possible (n = 54). The third training workshop combined messages from first and second training, emphasizing benefits and malleability of empathy (n=53). The fourth training workshop was a control or placebo workshop, enabling us to assess the impact of the training content independently of participating in any workshop on prosocial behavior (n = 53).

Experimental Details.—Each training workshop included a roughly 15-minute lecture and a structured discussion. After watching a 15-minute video lecture, participants completed a short writing assignment on the main lessons learned in the lecture and two weeks later participated in a structured discussion via a live Zoom session. The two-week interval between the lecture and discussion was motivated by the literature on social-emotional learning pedagogy, which suggests that spacing out doses over time can enhance learning (Walton and Cohen, 2011). Specifically, the structured discussion involved a recapitulation of the main lessons of the lecture video and with the following questions were asked from the junior ministers:

[4] Individual level randomization was performed using a random number generator in Stata.

[5] The script of the email sent out to all officers is presented in Table B2 in the Online Appendix B.

[6] We used the services of an expert computer scientist who blocked sharing and downloading of the training lecture.

"Q1. What do you think were the main messages of the lecture? Q2. How do you think you may apply lessons from the lecture to your job? Give at least 3 examples. The exact questions discussed to start the structured discussion can be found in Table B6 in Appendix B. Table B1 in Appendix B presents a flow chart of the timing and broader set-up of the experiment.

Utilitarian Treatment.— Our first treatment involved the participants watching a training lecture emphasizing the utility of empathy and how it can benefit them in their personal and professional life. The training reinforced this message by relying on two approaches: narratives and research studies, that is, both qualitative and quantitative evidence. The training lecture begins by a motivating example or a "puzzle": why profit maximizing firms like Google invest millions in training their employees in showing empathy, e.g., at the *Google Empathy Lab*, especially when it is costly for them. We then argue that this is a profit maximizing response on the part of Google. We build on this example and emphasize several (truthful) real-life stories of former deputy ministers who were known to be prosocial and empathic(as well as famous) for their stellar public service record. The training goes on to discuss the main findings of several studies that back up these narrative accounts. For instance, we discuss studies that show that demonstrating empathy benefits firms by making employees better able to deal with complex social relationships and hierarchies. The training also discusses studies showing how elite agents such as CEOs and senior managers are better able to motivate their employees, reduce shirking, and increase overall productivity and profits by displaying more empathy, especially towards their subordinates. The utilitarian training treatment concludes by reiterating the main message of this treatment: "*Qualitative and quantitative evidence backs the idea that showing empathy is good for you. It is not just the right thing to do but also the most sensible thing to do for your performance.*"[7] The complete script of the training is presented in Table B3 of Appendix B.

Malleability Treatment.— Our second treatment arm was provided with training emphasizing the malleability of empathy. That is, how empathy changes over time within a person and across populations. This treatment was inspired by prior work in psychology that documents that the degree of empathy a person has is not a fixed personality trait but is rather malleable. This literature finds that reminding subjects that empathy is not fixed can increase short-term empathic behavior (see Weisz and Zaki, 2017 for a review of this literature). The malleability training reinforced the malleability of empathy message by relying on two earlier approaches: qualitative and quantitative evidence. That is, this training relied on narratives of personal transformation – stories emphasizing the malleability of empathy – and quantitative research in

[7] The complete script of the training is presented in Table B3 of Appendix B.

psychology that argues that empathy is malleable and that people can become more prosocial over time. This focus on personal growth was reinforced via narratives and quantitative evidence. The malleability training also concludes by reinforcing the main message of this treatment: "*Qualitative and quantitative evidence backs the idea that empathy is not fixed but is malleable. It is a skill that can be developed.*"[8] In an effort to facilitate a clearer comparison of the treatment scripts, we have appended a color-coded transcript to this article. Specifically, passages that appear in both the Utilitarian (U) and Malleability (M) treatment scripts are marked in brown, text common to the Malleability and combined treatments is highlighted in green, and sections shared by the Utilitarian and combined treatments are denoted in blue. This color-coding system enables us to effectively differentiate and identify the content that is either shared or unique across the three distinct treatments, thereby providing a clear visual representation of their textual intersections and divergences.

Joint Utilitarian and Malleability Treatment.— Our third treatment arm received both utilitarian and malleability treatments together. This group was allocated the training that emphasized *both* the utility and malleability of empathy. Like our stand-alone treatments, this group received narrative accounts and quantitative evidence arguing that empathy is both beneficial for them and malleable. This training concludes by reinforcing the main message of this treatment: "*Qualitative and quantitative evidence backs the idea that empathy is good for you. It is not just the right thing to do but also the most sensible thing to do for your performance. Qualitative and quantitative evidence also backs the idea that empathy is not fixed but malleable. It is a skill that can be developed.*"The complete script for the joint utilitarian and malleability treatment is presented in Table B5 of Appendix B. At the bottom of the Table B5, we can find hyperlinks to the actual video and audio recordings of the treatments, accompanied by their respective durations. The Utilitarian treatment spans approximately 18 minutes, the Malleability treatment lasts around 13 minutes, and the duration of the combined treatment is also roughly 18 minutes.[9] The similar lengths of the combined and Utilitarian treatments suggest that differences in attention solely attributable to time are an unlikely factor in explaining our results.

Placebo.— Finally, our control group received a placebo training unrelated to the utility or malleability of empathy. They received a macroeconomics lecture taught in the economics undergraduate program at the Lahore School of Economics. The training lecture that this placebo group underwent covered basic

[8] The complete script for the training lecture treatment is presented in Table B4 of Appendix B.

[9] The lecture was initially anticipated to last for 30 minutes; which was stated in the start of the lecture. it was actually concluded in 18 minutes.

macroeconomic facts and concepts that include definitions and discussion of Gross Domestic Product, Gross National Product, Purchasing Power Parity and macroeconomic identities. All lectures, including the placebo, were delivered by the same person and every lecture ended with participants writing an essay summarizing key points of the lecture.

Balance.—Table 1, reports individual level summary statistics by treatment group. Differences across treatment groups are small in magnitude, and almost all p-values estimates are larger than 0.10, suggesting that the randomization was effective at creating balance between the groups. For instance age, gender, birth in political capitals, asset ownership, and foreign visits are balanced across randomly assigned groups.[10] Most salient to note are pre-treatment outcomes, in particular those related to altruism. From the top rows of Table 1, we observe that baseline blood donations and scores on pre-treatment psychological assessments used to screen antisocial candidates are also balanced. The groups are also balanced in pre-treatment measures of cognitive ability such as mathematics and written examination scores, as well as non-cognitive ability interview assessments. The similarity of *baseline blood donations*, and across pre-treatment written, mathematics, interview, and psychological assessments strongly suggest that the different treatment groups are balanced in both individual characteristics and pre-treatment altruism.

Book Roll out.— Three months after the initial intervention, we cross-randomized deputy ministers to either get an empathy book or a placebo book.[11] The empathy book is Mindsight: Transform Your Brain with the New Science of Empathy by Daniel J. Siegel, a popular cognitive psychology book that suggests ways to cultivate empathy. This cross-randomization was to reinforce the impact of empathy workshops. We reinforced the book training with 30-minute video lectures by the authors of the books they received. The ministers then write two 1500-word essays on the main lessons of the book. One essay summarized every chapter of their assigned book, and the second involved how the materials would apply to their career. The essays were graded and rated in a competitive manner among ministers with treated and placebo books. Winners received monetary vouchers and peer recognition via commemorative shields. Specifically, we announced the first three positions for both groups assigned the book and distributed the commemorative shields and gift vouchers to a luxury departmental store. Table A6 of Appendix A shows that the book intervention (by itself and in interaction with the utilitarian treatment) does not have statistically significant

[10] Following Duflo et al., 2015, Table 1 reports standard deviations in brackets and p-values corresponding to respective F-statistics in italics.

[11] The placebo book is "Mastering 'Metrics" by Angrist and Pischke (2014). The identification assumption is that econometrics does not influence empathy.

impacts on orphanage visits and volunteering. This suggests that the original utilitarian treatment plays an important role in the effects that we observe.

Discussion of Power.— The focus on deputy ministers that make high-impact policy decisions allows us to study an elite group of high-stakes decision-makers who can potentially impact long-run economic development. Nevertheless, the selective nature of these decision-makers indicates that they are by design few in number. Therefore, our sample is limited to about 200 deputy ministers, which raises concerns about lack of statistical power. Nevertheless, even with 200 individuals, our evidence complements several important experimental studies that inspired subsequent work. For instance, the Abecedarian Program (n = 111) (Muennig et al., 2011), the Perry Preschool Program (n = 123) (Heckman and Karapakula, 2019); and the Jamaican Study (n = 129) (Grantham-McGregor et al., 1991).

This study addresses a topic, altruism in governance, that inherently faces challenges due to a small sample size. To be transparent about statistical power of our analysis, we have calculated the minimum detectable effects (MDEs) for our primary outcomes, which include specific game measures and a behavioral indicator, willingness to donate blood. These MDEs were determined based on the control group's means and standard deviations. Our power calculation with statistical power 80% and significance level of 5% reveals that in our sample, the individual level randomization with 53 ministers to a treatment group, so 106 for any comparison between two, allows us to detect a minimum detectable effect ranging from 0.2 to 0.8 standard deviations; Appendix Figure A1 shows the outcome with the highest MDE (appointment to donate) to lowest MDE (teamwork assessments).

Buttressing the assumption that our intervention is powered to have long-term effects, edutainment interventions have been shown to work (Riley, 2019; Banerjee et al., 2019). Self-persuasion interventions have been shown to have long-term effects (Eigen and Listokin, 2012; Schwardmann, Tripodi, and van der Weele, 2022). One study found long-term reductions in IAT scores with a multi-faceted prejudice habit-breaking intervention; there was a reduction in implicit race bias by 0.46 in standard deviation (Devine et al. 2012).

We further investigate the effect size estimates derived from the Stand-alone Utilitarian (U) training for which we found a statistically significant effect. Utilizing the methodological framework advanced by Gelman and Carlin (2014), we calculate the probability of committing a Type S error, which pertains to the incorrect inference of the direction of an effect, as well as the likelihood of a Type M error, which involves the overestimation of the magnitude of an effect size. This is achieved by juxtaposing the estimated effect sizes and standard errors against a series of hypothetical true effect sizes, posited to be 100%, 75%, 50%,

25%, and 12.5% of the original estimates reported in our study. Through this analysis, reported in Table B15, we ascertain the extent to which our effect size estimates might be subject to potential inflation or directional miscalculation.[12] The results suggest both are unlikely. Second, we benchmark our effect sizes against recent experimental research with similar designs and challenges, such as smaller sample sizes. Table B16 outlines the effect sizes and minimum detectable effects (MDEs) from our critical results on interventions designed to alter perceptions and actions. Calculated with a 0.05 significance level and 80% power, our effect size (0.52 SD) aligns with these other studies (see Table B16). Third, while experimental design variances may impact effect sizes, our study and two others show treatment effects surpassing the MDEs, a contrast to most studies (see Figure B1). To summarize, our analysis demonstrates that despite the limitations of a smaller sample size, our study design and results are robust and comparable to other experimental research in the field, thus providing reliable conclusions.

II. Data and Empirical Strategy

A. The Data

The outcome variable data on behavioral games was collected during a Zoom call with everyone under the in a live session. All the officers participated in 12 behavioral games during the 2-hour workshop.

Outcome Variables on Altruism.— Our first set of measurements assesses altruism. The first outcome variable is the standard measure of altruism, i.e., response of participants in a "dictator" game. Pioneered by Kahneman et al. (1986), the decision of the "dictator" to voluntarily donate money without clear benefit is widely regarded as a prominent measure for altruism and applied in many studies in economics and psychology (see Engel, 2011 for a review of this literature).[13] We consider the decision of the dictator as our first measure of altruism and our choice is motivated by the game holding in many real world settings of altruistic behavior (Henrich et al., 2005; Levitt and List, 2007; Kosse et al., 2020).[14] Our setting of implementing the dictator game is also interesting since instead of playing these games with students that

[12] We adopt the framework proposed by Gelman and Carlin (2014) and implemented by Shem-Tov, Raphael and Skog (2021), to estimate the probability of sign error (Type S error) and the average potential exaggeration ratio (Type M error) in the main treatment effect estimates in the paper. We can see that for the significant estimates, a sign error is very unlikely to occur, and the overall potential exaggeration ratio (i.e., inflation) is around 1.2, which would not contradict the main conclusions. An exaggeration ratio of 1.2 means that the estimated coefficient is at most 20% larger than the true coefficient.

[13] Specifically, the dictator game is a variant of the ultimatum game where strategic concerns are absent as the proposer simply states what the split will be and there is no veto power to affect the proposal on part of the recipient (Güth et al., 1982).

[14] Although Henrich et al. (2005) note that "context matters" and that there is large variation in the exact degree of altruism demonstrated that depends on the prevalent social norms in the society.

have self-selected for the experiment, we administer these games with deputy ministers, complementing the important new work that moves beyond student populations (see e.g. Cappelen et al., 2015). The second outcome variable is another variant of the dictator game – the charity game (Bettinger and Slonim, 2006). Participants are given the option to donate money to UNICEF to buy an effective measles vaccine and were provided with the information that this vaccination is likely to save lives. However, the money could only be sent at the expense of forgoing some money for themselves. This is similar to many studies that combine the standard dictator game with this variant of a charitable donation decision to assess whether results hold in both instances (see, e.g., Sutter et al., 2019). The outcome variables of the behavioral games are normalized between 0 and 1 to make the comparisons across games easier. In Appendix B, we also present results for outcome variables standardized to mean zero and standard deviation one. Our third set of measurements assesses prosociality in the field. In collaboration with a volunteer group working for a prominent blood bank in Lahore, we designed and randomized the script for volunteers making the telephone calls on behalf of the blood bank to all deputy ministers with an urgent, but truthful, request to donate blood.[15] We measure outcomes for the public servants agreeing to donate blood as well as those actually agreeing to set up a definite appointment to donate blood at the blood bank.[16] The phone calls requesting blood donations took place about two months following the roll-out of our training lectures and submission of the summary. The remaining individuals within each treatment arm were randomly assigned an urgent generic request for blood donation without explicit mention of the blood type of the deputy ministers.[17] Besides donation of blood, we also measure donation of time. Two regular field trips took place about four months following the training. In the first field trip the policymaker must choose between attending a lecture by a senior bureaucrat or visiting an orphanage. In the second trip, the deputy minister must choose between volunteering in impoverished schools at a selected government network of schools or attending a lecture by a senior bureaucrat. In our average effect size analysis, we combine blood donations, orphanage visits, and volunteering in impoverished schools to create the field index of altruism.

[15] The urgency was truthful because the COVID-19 pandemic led to a steep fall in blood donations which created a shortage of all blood types . According to one of the volunteers making the calls: "the blood banks were practically empty".

[16] Both responses were recorded in the same phone call.

[17] Specifically, in the first group, a request is made to the deputy ministers that their blood type is urgently needed, for instance, "Blood for group O positive is urgently needed at the blood bank" (where the minister had O positive blood type), while the second group is requested to donate blood but without mention of the exact blood type of the bureaucrat, i.e., a generic request that "blood is urgently needed at the blood bank" is made.

Outcome Variables on Skill Assessments.— Other measures include grades on soft skills, teamwork and research method assessments workshops. The soft-skills workshop tests on material related to skills associated with social skills, perspective-taking, negotiations, leadership, and cooperation. The teamwork workshop is scored by a panel of senior bureaucrats, policymakers and academics and involves policy responses within a team. For instance, consider the sample scenario question, posed to the deputy ministers: "*The Prime Minister wants you to devote more resources to his security detail, while the Chief Minister wants you to aid in the flood relief efforts. How would you organize your team? What decisions will you take? Please detail the exact steps.*"

Sample Size and Randomization Inference. — Imbens and Rubin (2015) recommend — in small sample randomized trials — conducting randomization inference where the econometrician scrambles the data, reassigning treatments and comparing the distribution of placebo estimates with the true estimate from the experiment.[18] We report in Table B14 of Appendix B the corresponding p-values with 1000 iterations of this process applying the most strict criteria of nesting all 36 outcomes in a single family. Even though the p-values slightly increase, the treatment effects are still statistically significant at conventional levels. These results strongly suggest that idiosyncratic small sample bias is unlikely to explain our results. Supporting this conclusion is the fact that lasso-selected controls do not affect the robustness of the results (see, e.g. Appendix A Table A1 and A3).

B. Attrition

Close cooperation with the partner implied that we had 100% take-up of our treatments. There was, nevertheless, some attrition in recording our blood donations outcome variable in the field. Roughly 95% of participants gave definite responses to both the blood donation requests and setting up a definite appointment with the blood bank. We do, however, show that, even with this small dropout rate, there is no evidence for differential attrition for both agreeing to donate blood or setting up a definite appointment for the blood donation (these results are reported in Table B7 of Appendix B).

C. *Estimation Strategy*

[18] *ritest* in Stata is implemented to compute p-values corresponding to the permutation inference. The results are robust to choosing different numbers of iterations.

The impact of the two stand-alone utilitarian and malleability training and the joint training can be evaluated by comparing outcomes across groups in a simple regression framework. For each individual-level outcome, the estimation equation is:

$$Y_i = \alpha + \beta U_i + \gamma M_i + \delta UM_i + X_i \mu + \epsilon_i \quad (1)$$

where Y_i is respective outcome for deputy minister i, U_i is a dummy equal to one if the deputy minister is assigned to the stand-alone utilitarian empathy treatment arm; M_i is a dummy variable equal to one if the deputy minister is assigned to the stand-alone malleability empathy treatment arm; UM_i is a dummy variable equal to one if the deputy minister is assigned to the joint utilitarian and malleability treatment arm; X_i is a vector of individual-level controls. We cluster standard errors at the individual level since that is our level of randomization. In equation (1), β measures the effect of stand-alone utilitarian treatment; γ the effect of stand-alone malleability treatment; and δ the effect of the joint treatment. In all tables that follow, we present estimates of equation (1) for a series of outcomes. At the bottom of each panel, we show the mean of the dependent variable for the placebo group, and we present p-values for tests of the hypothesis that the effect of the joint treatment is equal to either of the two stand-alone treatments, or equal to the sum of the two stand-alone treatments (i.e we test for $\beta = \gamma$, $\gamma = \delta$ and $\delta = \beta + \gamma$). We report ordinary least squares (OLS) estimations. The results are qualitatively unchanged with probit or logit estimations for binary outcomes.

Explanatory Variables.— Our main treatment variables are dummies for the three treatments. U_i and M_i are dummies that switch on if an individual deputy minister is assigned to the stand-alone utilitarian, stand-alone malleability and UM_i joint utilitarian and malleability treatment arms, respectively. We add as control variables all the individual characteristics available from administrative data. These individual level control variables are as follows: written, mathematics, psychological and interview assessment scores in entry examination; income before joining the service; age; years of education and dummies for gender, birth in political capitals, asset ownership, foreign visits and occupational or professional designation.

III. Results

A. *Impact on Altruism*

Columns (1) and (2) of Table 2 present the estimated effects of our three treatments relative to the placebo group in the classic dictator game. We find that only the stand-alone utilitarian treatment significantly

increases altruism. Since we have normalized the outcome variable to be between 0 and 1, we can infer that the utilitarian treatment increases altruism by about 6 percentage points. This is equivalent to a 12% increase over the placebo mean. The coefficient estimates are similar with no controls and a large number of individual level characteristics added in the regression. Likewise, in Table 2, we also report results of a variant of the dictator game when donations to UNICEF charity are solicited instead of donations to strangers as in the previous standard dictator game. The effects are even larger and reported in columns (3) and (4) of Table 2: utilitarian treatment is associated with a 20 percentage point increase in altruism scores, or a 33% increase over the placebo mean. Equivalently, the utilitarian treatment increases altruism in dictator and charity games by about 0.3 to 0.5 standard deviation relative to the placebo group. These results are also reported in Table B8 of Appendix B where we standardized the outcome variables to mean zero and standard deviation one. Table A1 and A3 in Appendix A present similar results with Lasso controls, while Table A7 and A8 report the results with standardization done with respect to the placebo group. For comparison, the effect sizes of our utilitarian training intervention (video lecture, summary and book receipt) are about as large as the effect found from a year-long mentoring program aimed at enhancing "other-regarding behavior" in 7–9 year olds in Germany (Kosse et al., 2020). These results are corroborated by evaluation of a regular soft-skills assessment.

B. *Field Evidence from Blood Donations and Orphanage Visits*

We leverage unique information on the blood groups of the deputy ministers and randomized phone calls to provide results from the field. In collaboration with a prominent blood bank, we randomized the phone calls to the deputy ministers so that half of them (106 participants) were randomly told that their particular blood group was in urgent need, while the other half (107 participants) were just provided with an urgent request to donate blood but without any mention that their exact blood group was needed. That is, the first group gets the call "O Positive blood is urgently needed" (where the deputy minister had the O Positive blood group), whereas the second group gets a generic request that "Blood is urgently needed". The first two columns of Table 3 report the results on agreement to donate blood, while the latter two columns report results on responses on setting up a definite appointment to donate blood at the bank. The estimates presented in Columns (1) and (3) reveal a large effect of the utilitarian treatment: the stand-alone utilitarian group is about 25 percentage points more likely to both agree to donate blood and set up a definite appointment with the blood bank relative to the group that received the placebo training. This is a substantial effect and equivalent to about 80% increase over the placebo mean. These results are also reported as a bar

chart in Figure 2: the group assigned stand-alone utilitarian treatment has about 25 percentage points higher blood donations relative to the placebo group on both blood donation variables (Figure 2, Panel A and B). This strongly suggests that results from behavioral games map well to real-life altruistic behavior in the field. Only the stand-alone utilitarian treatment has a qualitatively and statistically different effect on blood donations relative to the placebo group, consistent with the results from dictator games and empathy book choice. However, this doubling of blood donations for the group assigned the utilitarian treatment masks important heterogeneity among those that were randomized into the group that was told that their exact blood group was in need, relative to those that were made a generic request to donate blood. Columns (2) and (4) of Table 3 report estimates on the interaction terms of the three treatments with the randomly assigned status of the blood bank requesting the minister's actual blood type for both blood donation variables. Remarkably, the effect of blood donations seems to be *entirely* explained by the utilitarian group when the blood bank requested that their exact blood type was needed. These results can be observed most clearly in Figure 2: we observe that the blood donations more than doubled for the utilitarian group when their matching blood type was requested (left panel). We, nevertheless, do not find any significant difference in blood donations between utilitarian and placebo groups when the generic requests for blood donations were made (right panel). The deputy ministers who were assigned the utilitarian treatment are only willing to donate blood if their exact blood group is requested.

Additional evidence corroborates the view that the utilitarian group displays greater altruism in the field. These data are collected separately from the research team and unlikely to be affected by experimenter demand effects. Consistent with the results on blood donations, we find that the group assigned the stand-alone utilitarian treatment is about 20 percentage points more likely to make field visits to the orphanage relative to attending the lecture from the policy official (Table 3, Column 5). This is equivalent to about an 80% increase over the placebo mean. These results are corroborated with a second field trip six months after the treatment and two months after the orphanage visits: the deputy ministers have the choice between volunteering to teach for a week in any impoverished government school that falls under the Progressive Network or once again choose to attend a lecture on government programs from a senior public official. We also find that the group assigned the stand-alone utilitarian treatment is about 20 percentage points more likely to volunteer at impoverished schools. Substantively, the results on "field trips" are interesting for two key reasons: (1) the field visits and volunteering at impoverished schools took place at the end of January, that is, about four months after our trainings, and (2) these data come directly from the regular training curriculum, providing an external corroboration of our results.

C. *Performance in Assessments*

To corroborate our results to measure outcomes beyond those designed by us, we use assessment scores in tests. This includes assessments in teamwork, soft-skills and quantitative research methods. The teamwork assessment is used to gauge their group performance as junior ministers. Teamwork assessment in group tasks and tests ministers in teambuilding and leadership in policy situations. Effective teamwork is a likely consequence of soft skills as noted in Deming and Weidmann (2021) who have shown in important new work that soft skills are key to teamwork in the laboratory. We also have available a soft-skills assessment and a quantitative research methods assessment. The soft skills assessment tests ministers on negotiation, social skills, and cooperation in policy scenarios, while the research methods assessment tests them on hypothesis testing, multivariate regression analysis with applications to policy-making, and randomized evaluations. The research methods assessment serves as a placebo since it is not directly related to altruism or soft-skills. Columns (1) and (2) of Table 4 present these results: we observe that stand-alone utilitarian treated ministers have about 0.5 standard deviation higher scores in their teamwork policy assessments relative to the placebo group, while we find no evidence of malleability or joint treatment significantly impacting these team assessment scores. Columns (3) and (4) report scores on the soft-skills exams and also find elevated levels of assessment scores for the group assigned the utilitarian value of empathy treatment. In contrast, we find no impact on quantitative research method courses (Table 4, Columns 5 and 6). These results strongly suggest that the utilitarian treatment has a real impact on soft skills.

D. *Behavioral Evidence of Perspective-Taking*

The results so far show training policymakers in the benefits of empathy increases altruism, teamwork, and field outcomes related to successful mentalizing relevant to thinking of others. Here, we show that the impacts of training the utility of empathy extend to measurements traditionally utilized in laboratory settings to proxy for soft-skills. Table 5 presents estimates of the impact of our treatment in cooperation and coordination (Sutter et al., 2019). In the cooperation game, a decision maker must decide how much of an endowment to transfer to the other participant. The transferred quantity will be doubled and the other participant will receive this doubled quantity. What is not transferred remains in the decision maker's possession and is not doubled. At the same time, the other participant simultaneously makes the same decisions. This game is intended to reflect real-world situations where people must cooperate to achieve higher joint surplus. In the coordination game, the person chooses between two options. If the decision

maker and the other participant both choose one of the options, they will both receive higher joint surplus, which is split equally. However, there is an incentive to deviate, which is also the safe option that guarantees a non-zero outcome for the decision maker. This game is intended to reflect real-world situations where people must coordinate in teams. Several studies suggest related games map well into behavior in real-world teams (Grossman and Baldassarri, 2012; Barr and Serneels, 2009).

In Table 5 Columns (1) and (2), we observe individuals receiving the stand-alone utilitarian treatment perform better in the cooperative decision-making behavioral game. Specifically, they score 14 percentage points higher in this game than the placebo group. Likewise, in Columns (3) and (4), we find that these public officials also perform better in the coordination game: the group receiving stand-alone utilitarian treatment have about 7 percentage points higher scores in the Nash equilibrium coordination game. Equivalently, the deputy ministers assigned the utilitarian treatment arm score 0.4 of a standard deviation higher in decision-making and coordination.[19] Importantly, this suggests that cooperation and coordination, rather than simply redistributive preferences, drive the behavioral changes. This is relevant since high-skilled, "cognitive" occupations are increasingly valuing soft skills surrounding teamwork to enhance productivity (Deming, 2017). These results are consistent with successful mentalizing as in the case of increased blood donations when the decision makers were requested their exact blood type.

Honest public officials are also likely important for effective governance. Taking a long view, training altruism may increase prosociality by increasing honesty. This may have consequences among civil servants by making them more honest. The final game measures lying: each player rolls a 6-sided dice and is asked to report the outcome of the roll, but the player who reports a higher outcome also receives a higher payoff. There is an incentive to lie rather than truthfully revealing the die roll. That is, the public officials have the option of winning dishonestly by misreporting (see Fischbacher, et al., 2013; Hanna and Wang, 2017; Gneezy et al., 2018; Barfort et al. 2019). Figure 3 presents the results of the lying game. We find, remarkably, that the utilitarian group is significantly less likely to lie in the dice game relative to the placebo group. Interestingly, the stand-alone utilitarian group average is extremely close to 3.5 which is what would be obtained if everyone honestly revealed their truthful die-roll.

While we hypothesize that the successful mentalizing of others plays a key role for our results on altruism, we investigate and rule out alternative channels such as redistributive preferences or competitiveness. Namely, the results indicate altruism, not just fairness; effective altruism, not just altruism (because blood

[19] The standardized equivalent to Table 5 where dependent variables are standardized to mean zero and standard deviation 1 can be found in Table B9 of Appendix B.

donations increase only when they know that the decision to donate is more likely to be useful); and learning, not just priming or experimenter demand effects. For instance, the utilitarian treated group may have become more competitive, donating blood as a way to compete with their peers. This would be consistent with the fact that the utilitarian training lecture emphasized that showing empathy is a utility maximizing response. If that were the case, we should see blood donations increasing regardless of their explicit blood type being requested. Alternatively, one could reason that the utilitarian treatment made the public officials more redistributive, or patient, or trusting and this is what explains the result in altruism games and blood donations in the field. Nevertheless, we do not find much evidence of this in the other behavioral games that the deputy ministers played. Table 6 reports these results.[20] We find no effect of any of our treatments on competitiveness, patience, perseverance, redistribution, risk and trust games (these games are discussed in Berg et al., 1995; Fisman et al., 2007; Bartling and Fischbacher, 2012; Dohmen et al., 2018; Bašić et al., 2020; Falk et al., 2020).

This exploratory analysis of mechanisms is also summarized in Figure 4, where we depict the estimated standardized (mean zero standard deviation one) stand-alone utilitarian treatment effects and 95 percent confidence intervals on coordination, cooperation, honesty, competitiveness, patience, perseverance, redistribution, risk aversion and trust games. The thing that stands out in this picture is that coordination, perspective-taking and honesty are likely to be a common mechanism responsible for the treatment effects we estimate, while changes in patience, perseverance, redistribution, risk preferences or trust are unlikely to be driving the results. Therefore, the data consistently suggest that treated junior ministers are more inclined towards altruistic actions than the control group. They are not only more likely to donate blood when it is most needed but also demonstrate improved coordination and cooperation. Finally, we summarize all the results in Appendix Table B12, we show the results by combining our outcome variables as a composite index of Altruism based on the Average Effect Size (AES) approach of Kling et al., 2004.[21]

IV. Conclusion

[20] The null results are essentially identical if we standardize the dependent variable instead of normalizing it. See Table B10 in Appendix B for these results.

[21] We also perform a randomization inference check in Table B13 and conduct a MHT robustness check, where we adjust for the fact that we are testing for multiple hypotheses by using sharpened False Discovery Rate (FDR) q-values. Similar results are obtained when we deploy List et al., (2019)'s familywise error rate correction (FWER); this extends the False Discovery Rate (FDR) method by using a bootstrapping approach, incorporating point-dependence structure of different treatments and controlling for the familywise error rate i.e., the probability of one or more false rejections (see Table B14 of Appendix B).

We find that training high-stakes decision makers in different schools of thought to cultivate prosociality yielded significant impacts from training in the utility of empathy. Soft-skills have been formally modeled to reduce coordination costs so that teams, organizations, and society work together more effectively. We provide causal evidence on the impacts of training utility of empathy on soft-skills of deputy ministers' teamwork and coordination that is critical in models of soft skills.

Laboratory measures of altruism, charitable donations, cooperation and coordination were impacted. Independent assessments of teamwork and skills as deputy ministers also increased. Treated ministers doubled their blood donations in response to blood banks—but only when the specific blood type matching the minister was requested. Orphanage visits and volunteering also increased. In terms of effect sizes, training the utility of empathy has a similar effect size on prosocial behavior (0.4-0.6 standard deviation) as a one-year mentoring program of elementary school children (Falk et al., 2020).

It is unlikely that experimenter demand effects drive our results – i.e., deputy ministers in the utilitarian treatment behaving in a way they feel they are expected to by the experimenter. This is due to several reasons. First, the treatment group only responded to blood bank donation requests when their exact blood type was requested. Second, malleability also emphasized empathy, and experimenter demand effects would plausibly also affect those treatment groups as well. Third, a number of high-stakes administrative assessment scores including soft-skills and teamwork assessments were conducted separately from the research team.[22] The measurements and patterns in the data, therefore, indicate that experimenter demand effects are unlikely to explain our results. Taken together, our sensitivity analysis strongly suggests that our results are robust to multiple hypothesis testing, experimenter demand, small samples, and lack of balance on utilitarian treatment impact on prosocial behavior.

This research explores the concept of altruism within governance, a topic that is notably challenging due to limited sample sizes. To maintain transparency regarding the statistical power of our findings, we have calculated the minimum detectable effects (MDEs) for key outcomes. We view these results as a WAVE1 insight, in the nomenclature of List (2020), and replications need to be completed to understand if the effect sizes can be applied to other general populations as well as high-stakes decision makers in other contexts. Following List (2020)'s SANS (Selection-Attrition-Naturalness-Scaling) conditions for generalizability of our results, we offer three comments. First, in terms of selection, our sample consists of all 213 elite policy makers that entered service in Pakistan via competitive examinations in a given year. Second, our

[22] We also observe no impact of the malleability treatment on prosocial behavior which is also inconsistent with experimenter demand effects explaining our results.

compliance is nearly 100% in the behavioral games as they were held in the natural setting, while in blood donations, volunteering and orphanage visits we still have close to 90% compliance given the credibility of prominent blood bank soliciting calls. The setting and choice tasks are natural measures. The policy makers in their field decisions and test assessments are not placed on an artificial margin, rather, they are performing natural tasks in the field. Third, in terms of scaling our intervention to increase altruism in other settings, the intervention is cheap to deliver, parsimonious, and may be particularly useful for developing countries who face strict resource constraints.

Much attention has focused on childhood interventions, though some work on workplace-based programs that teach character skills have made important strides, yet no randomized control trial attempts to train prosociality in different schools of thought in adults (Kautz et al., 2014). We show that empathy can be enhanced even among adults (Barrera-Osorio et al., 2020 and Chioda et al., 2021) which is consistent with the evidence that cognitive behavioral therapy impacted outcomes of adults in Liberia (Blattman et al., 2017). Future research could test additional schools of thought that offer a parsimonious foundation for normative ethics besides the two in our study and investigate their welfare consequences.

References

1. Abeler, J., Nosenzo, D. and Raymond, C., 2019. Preferences for truth-telling. *Econometrica, 87*(4), pp.1115-1153.
2. Alan, S., Baysan, C., Gumren, M. and Kubilay, E., 2021. Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *The Quarterly Journal of Economics, 136*(4), pp.2147-2194.
3. Alan, S. and Ertac, S., 2018. Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy, 126*(5), pp.1865-1911.
4. Alger, I. and Weibull, J.W., 2013. Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica, 81*(6), pp.2269-2302.
5. Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103*(484), pp.1481-1495.
6. Angrist, J.D. and Pischke, J.S., 2014. *Mastering'metrics: The path from cause to effect*. Princeton university press.
7. Ashraf, N., Bandiera, O., Davenport, E. and Lee, S.S., 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review, 110*(5), pp.1355-94.
8. Autor, D.H., Dorn, D. and Hanson, G.H., 2015. Untangling trade and technology: Evidence from local labour markets. *The Economic Journal, 125*(584), pp.621-646.
9. Banerjee, A., Pasea, L., Harris, S., Gonzalez-Izquierdo, A., Torralbo, A., Shallcross, L., Noursadeghi, M., Pillay, D., Sebire, N., Holmes, C. and Pagel, C., 2020. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *The Lancet, 395*(10238), pp.1715-1725.
10. Ballesta, S., Shi, W., Conen, K.E. and Padoa-Schioppa, C., 2020. Values encoded in orbitofrontal cortex are causally related to economic choices. *Nature, 588*(7838), pp.450-453.

11. Barfort, S., Harmon, N.A., Hjorth, F. and Olsen, A.L., 2019. Sustaining honesty in public service: The role of selection. *American Economic Journal: Economic Policy, 11*(4), pp.96-123.
12. Bartling, B. and Fischbacher, U., 2012. Shifting the blame: On delegation and responsibility. *The Review of Economic Studies, 79*(1), pp.67-87.
13. Barr, A. and Serneels, P., 2009. Reciprocity in the workplace. *Experimental Economics*, *12*(1), pp.99-112.
14. Barrera-Osorio, F., Kugler, A. D., & Silliman, M. I. (2020). Hard and soft skills in vocational training: Experimental evidence from Colombia (No. w27548). National Bureau of Economic Research.
15. Bašić, Z., Falk, A. and Kosse, F., 2020. The development of egalitarian norm enforcement in childhood and adolescence. *Journal of Economic Behavior & Organization, 179*, pp.667-680.
16. Batson, C.D., 1997. Self–other merging and the empathy–altruism hypothesis: Reply to Neuberg et al.(1997).
17. Bénabou, R. and Tirole, J., 2004. Willpower and personal rules. *Journal of Political Economy, 112*(4), pp.848-886.
18. Bénabou, R. and Tirole, J., 2006. Incentives and prosocial behavior. *American economic review, 96*(5), pp.1652-1678.
19. Bénabou, R. and Tirole, J., 2011. Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics, 126*(2), pp.805-855.
20. Berg, J., Dickhaut, J. and McCabe, K., 1995. Trust, reciprocity, and social history. *Games and economic behavior, 10*(1), pp.122-142.
21. Besley, T. and Ghatak, M., 2018. Prosocial motivation and incentives. *Annual Review of Economics*, *10*, pp.411-438.
22. Bettinger, E. and Slonim, R., 2006. Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics, 90*(8-9), pp.1625-1648.
23. Blattman, C., Jamison, J.C. and Sheridan, M., 2017. Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia. *American Economic Review, 107*(4), pp.1165-1206.

24. Bloom, N. and Van Reenen, J., 2011. Human resource management and productivity. In *Handbook of labor economics* (Vol. 4, pp. 1697-1767). Elsevier.
25. Burks, S.V., Nosenzo, D., Anderson, J., Bombyk, M., Ganzhorn, D., Götte, L. and Rustichini, A., 2016. Lab measures of other-regarding preferences can predict some related on-the-job behavior: Evidence from a large scale field experiment.
26. Cappelen, A.W., Nygaard, K., Sørensen, E.Ø. and Tungodden, B., 2015. Social preferences in the lab: A comparison of students and a representative population. *The Scandinavian Journal of Economics*, *117*(4), pp.1306-1326.
27. Cappelen, A., List, J., Samek, A. and Tungodden, B., 2020. The effect of early-childhood education on social preferences. *Journal of Political Economy*, *128*(7), pp.2739-2758.
28. Chen D.L., Schonger M., Wickens C., 2016. oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, pp. 88-97.
29. Chioda, L., Contreras-Loya, D., Gertler, P., & Carney, D. (2021). *Making Entrepreneurs: Returns to Training Youth in Hard Versus Soft Business Skills* (No. w28845). National Bureau of Economic Research.
30. Clingingsmith, D., Khwaja, A.I. and Kremer, M., 2009. Estimating the impact of the Hajj: religion and tolerance in Islam's global gathering. *The Quarterly Journal of Economics*, *124*(3), pp.1133-1170.
31. Cooper, D.J. and Kagel, J.H., 2016. 4. Other-Regarding Preferences. In *The Handbook of Experimental Economics, Volume 2* (pp. 217-289). Princeton University Press.
32. Deming, D.J., 2017. The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), pp.1593-1640.
33. Weidmann, B. and Deming, D.J., 2021. Team players: How social skills improve team performance. Econometrica, 89(6), pp.2637-2657.
34. Dohmen, T., Falk, A., Huffman, D. and Sunde, U., 2018. On the relationship between cognitive ability and risk preference. *Journal of Economic Perspectives*, *32*(2), pp.115-34.
35. Duffau, H., 2014. The huge plastic potential of adult brain and the role of connectomics: new insights provided by serial mappings in glioma surgery. *Cortex*, *58*, pp.325-337.

36. Duflo, E., Dupas, P. and Kremer, M., 2015. Education, HIV, and early fertility: Experimental evidence from Kenya. *American Economic Review, 105*(9), pp.2757-97.
37. Eisenberg, N., Miller, P.A., Shell, R., Mcnalley, S. and Shea, C., 1991. Prosocial development in adolescence: A longitudinal study. *Developmental psychology, 27*(5), p.849.
38. Eisenberg, N. and Miller, P.A., 1987. The relation of empathy to prosocial and related behaviors. *Psychological bulletin, 101*(1), p.91.
39. Engel, C., 2011. Dictator games: A meta study. *Experimental economics, 14*(4), pp.583-610.
40. Falk, A., Neuber, T. and Szech, N., 2020. Diffusion of being pivotal and immoral outcomes. *The Review of Economic Studies, 87*(5), pp.2205-2229.
41. Fehr, E. and Gächter, S., 2002. Altruistic punishment in humans. *Nature, 415*(6868), pp.137-140.
42. Fischbacher, U. and Föllmi-Heusi, F., 2013. Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association, 11*(3), pp.525-547.
43. Fisman, R., Kariv, S. and Markovits, D., 2007. Individual preferences for giving. *American Economic Review, 97*(5), pp.1858-1876.
44. Francois, P., Fujiwara, T. and Van Ypersele, T., 2018. The origins of human prosociality: Cultural group selection in the workplace and the laboratory. *Science advances, 4*(9), p.eaat2201.
45. Galinsky, A.D. and Moskowitz, G.B., 2000. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology, 78*(4), p.708.
46. Galinsky, A.D. and Ku, G., 2004. The effects of perspective-taking on prejudice: The moderating role of self-evaluation. *Personality and Social Psychology Bulletin, 30*(5), pp.594-604.
47. Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S.M. and Grantham-McGregor, S., 2014. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science, 344*(6187), pp.998-1001.
48. Gneezy, U., Kajackaite, A. and Sobel, J., 2018. Lying aversion and the size of the lie. *American Economic Review, 108*(2), pp.419-53.

49. Grantham-McGregor, S.M., Powell, C.A., Walker, S.P. and Himes, J.H., 1991. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *The Lancet*, *338*(8758), pp.1-5.
50. Grossman, G. and Baldassarri, D., 2012. The impact of elections on cooperation: Evidence from a lab-in-the-field experiment in Uganda. *American journal of political science*, *56*(4), pp.964-985.
51. Guiso, L., Sapienza, P. and Zingales, L., 2009. Cultural biases in economic exchange?. *The quarterly journal of economics*, *124*(3), pp.1095-1131.
52. Guardian, 2000. The most dangerous man in the world. Retrieved from: https://www.theguardian.com/lifeandstyle/1999/nov/06/weekend.kevintoolis.
53. Gulzar, S. and Khan, M.Y., 2021. 'Good Politicians': Experimental Evidence on Motivations for Political Candidacy and Government Performance. Mimeo.
54. Güth, W., Schmittberger, R. and Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, *3*(4), pp.367-388.
55. Hanna, R. and Wang, S.Y., 2017. Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, *9*(3), pp.262-90.
56. Heckman, J., Pinto, R. and Savelyev, P., 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), pp.2052-86.
57. Heckman, J.J., García, J.L., and Ziff, A.L., 2018. Gender differences in the benefits of an influential early childhood program. *European economic review*, *109*, pp.9-22.
58. Heckman, J.J. and Karapakula, G., 2019. *The Perry Preschoolers at late midlife: A study in design-specific inference* (No. w25888). National Bureau of Economic Research.
59. Heller, S.B., Shah, A.K., Guryan, J., Ludwig, J., Mullainathan, S. and Pollack, H.A., 2017. Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics*, *132*(1), pp.1-54.
60. Henrich, J.P., Boyd, R., Bowles, S., Fehr, E., Camerer, C. and Gintis, H. eds., 2004. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press on Demand.
61. Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J. and Henrich, N.S., 2005. "Economic man" in cross-

cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, *28*(6), pp.795-815.

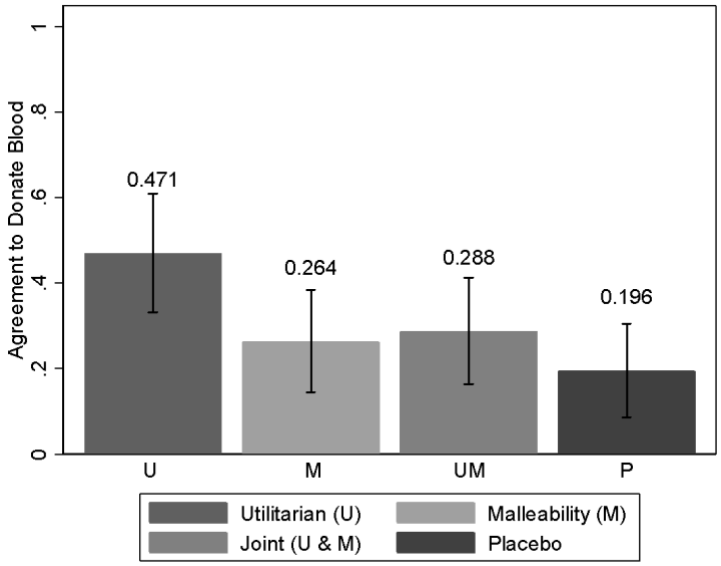
62. Imbens, G.W. and Rubin, D.B., 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.
63. Iyer, L. and Mani, A., 2012. Traveling agents: political change and bureaucratic turnover in India. *Review of Economics and Statistics*, *94*(3), pp.723-739.
64. Kahneman, D., Knetsch, J.L. and Thaler, R.H., 1986. Fairness and the assumptions of economics. *Journal of business*, pp.S285-S300.
65. Kautz, T., Heckman, J.J., Diris, R., Ter Weel, B. and Borghans, L., 2014. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. NBER Working Paper.
66. Kling, J.R., Liebman, J.B., Katz, L.F. and Sanbonmatsu, L., 2004. Moving to opportunity and tranquility: Neighborhood effects on adult economic self-sufficiency and health from a randomized housing voucher experiment. *Available at SSRN 588942*.
67. Knack, S. and Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of economics*, *112*(4), pp.1251-1288.
68. Kosse, F. and Tincani, M.M., 2020. Prosociality predicts labor market success around the world. *Nature communications*, *11*(1), pp.1-6.
69. Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H. and Falk, A., 2020. The formation of prosociality: causal evidence on the role of social environment. *Journal of Political Economy*, *128*(2), pp.434-467.
70. Levitt, S.D. and List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic perspectives*, *21*(2), pp.153-174.
71. Lindauer, M., Mayorga, M., Greene, J.D., Slovic, P., Västfjäll, D. and Singer, P., 2020. Comparing the effect of rational and emotional appeals on donation behavior. *Judgment and Decision making*, *15*(3).
72. List, J.A., Shaikh, A.M. and Xu, Y., 2019. Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), pp.773-793.
73. List, J.A., 2020. *Non est disputandum de generalizability? a glimpse into the external validity trial* (No. w27535). National Bureau of Economic Research.

74. Lopez-de-Silanes, F., La Porta, R.F., Shleifer, A. and Vishny, R.W., 1997. Trust in large organizations. *The American Economic Review*, *87*(2), pp.333-338.
75. Muennig, P., Robertson, D., Johnson, G., Campbell, F., Pungello, E.P. and Neidell, M., 2011. The effect of an early education program on adult health: the Carolina Abecedarian Project randomized controlled trial. *American journal of public health*, *101*(3), pp.512-516.
76. Ostrom, E.E., Dietz, T.E., Dolšak, N.E., Stern, P.C., Stonich, S.E. and Weber, E.U., 2002. *The drama of the commons.* National Academy Press.
77. Premack, D. and Woodruff, G., 1978. Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), pp.515-526.
78. Robalino, N. and Robson, A., 2016. The evolution of strategic sophistication. *American Economic Review*, *106*(4), pp.1046-72.
79. Shaukat Khanam Hospital, 2021. Shaukat Khanam Memorial Hospital and Research Centre. Blood Donations Club. Lahore.
80. Siegel, D.J., 2011. *Mindsight: The new science of empathy*.
81. Smith, A., 1759. *The theory of moral sentiments*. Alexander Kincaid and J. Bell, Edinburgh.
82. Smith, A., 1776. *The wealth of nations* (Vol. 11937). W. Strahan and T. Cadell, London.
83. Sutter, M., Zoller, C. and Glätzle-Rützler, D., 2019. Economic behavior of children and adolescents–A first survey of experimental economics results. *European Economic Review*, *111*, pp.98-121.
84. Walton, G.M. and Cohen, G.L., 2011. A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, *331*(6023), pp.1447-1451.
85. Weisz, E. and Zaki, J., 2017. Empathy building interventions: A review of existing work and suggestions for future directions. *The Oxford handbook of compassion science*, pp.205-217.

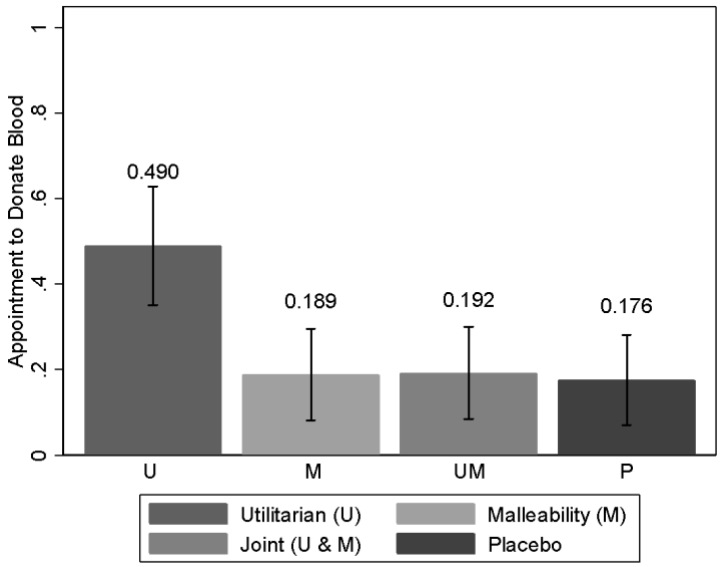
Figures and Tables

Figure 1: Impact on Blood Donations

Panel A: Agreement to Donate Blood

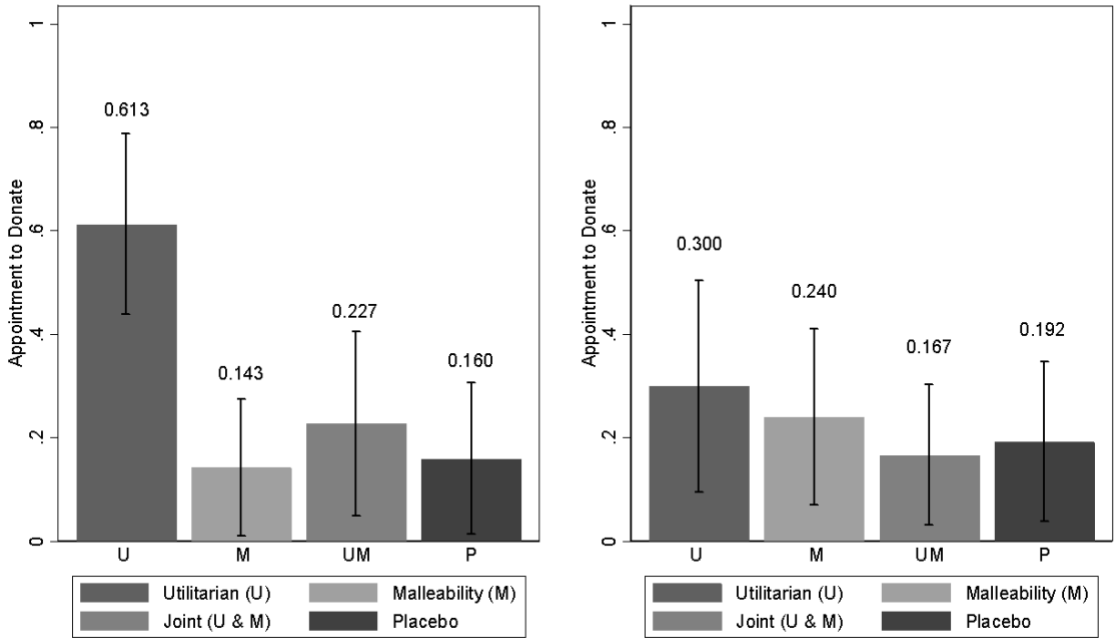


Panel B: Appointment to Donate Blood



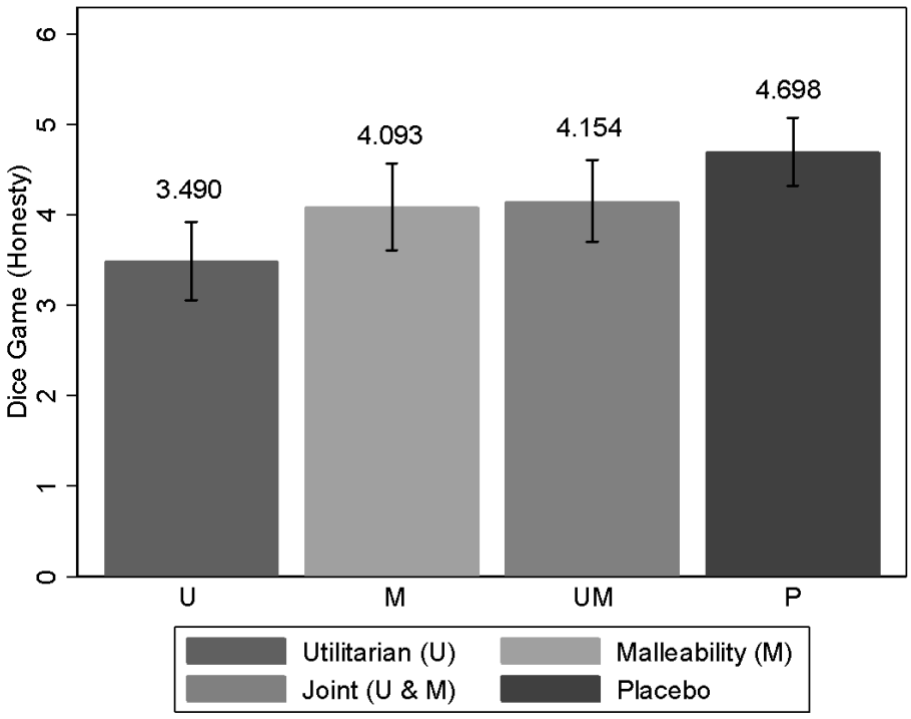
Note: The figure provides averages for the four randomly assigned groups along with 95% confidence intervals. Panel A provides averages for answers on the question of agreement to donate blood where one is yes, and no is zero. Likewise, Panel B provides averages for answers on setting an appointment with the blood bank to donate blood where yes is coded as one and no as zero.

Figure 2: Impact on Blood Donations by specific versus generic request



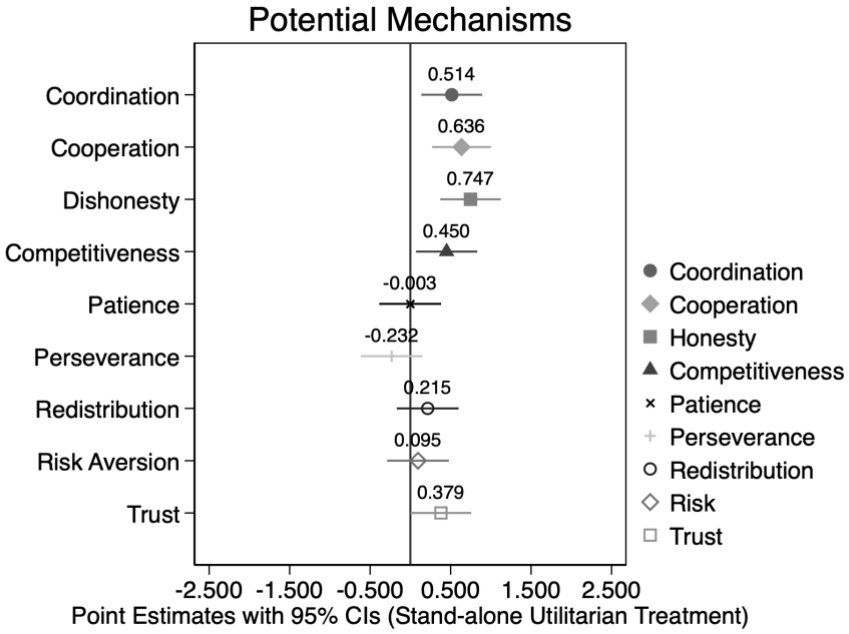
Note: The figure above provides averages for the four randomly assigned groups along with 95% confidence intervals. The figure on the left presents results on urgent truthful requests to donate blood with specific matching blood type of the individual, i.e., "O Positive Blood is urgently needed" (where the individual had the O Positive blood group). The figure on the right reports results from a generic request to donate blood i.e. "Blood is urgently needed".

Figure 3: **Effect on Lying Game**



Note: The figure provides averages for the four randomly assigned groups along with 95% confidence intervals. Each bar reports the average in the dice game. Higher levels represent more lying or dishonesty.

Figure 4: **Exploration of Mechanisms**



2

Note: The figure depicts the stand-alone utilitarian treatment effects and their 95% confidence intervals. Confidence intervals are based on standard errors. The vertical line indicates a treatment effect of zero. Dependent variables are standardized to mean zero and standard deviation one. Identical controls as in baseline specification are also always added.

Table 1: Baseline Characteristics, by Treatment Group

					Balance tests: p-value for test that:			
	Utilitarianism (U)	Malleability (M)	Utilitarianism & Malleability (UM)	Placebo (P)	U=P	M=P	UM=P	UM=U UM=M
Baseline Blood Donations	0.528	0.593	0.472	0.453	0.782	0.171	0.325	0.440
	[0.504]	[0.496]	[0.504]	[0.503]				0.151
Psychological Assessment Scores	7.302	7.167	7.283	7.302	*0.768*	*0.379*	*0.768*	*0.999*
	[1.085]	[1.240]	[0.968]	[1.137]				*0.475*
Writing Assessment Scores	653.802	651.480	660.401	656.735	*0.640*	*0.276*	*0.208*	*0.291*
	[36.224]	[28.718]	[36.377]	[29.999]				*0.152*
Interview Assessment Scores	132.788	129.360	131.623	130.600	*0.475*	*0.464*	*0.833*	*0.758*
	[24.272]	[18.591]	[21.760]	[16.800]				*0.566*
Math Assessment Scores	7.189	7.259	7.019	7.415	*0.817*	*0.883*	*0.184*	*0.502*
	[1.039]	[1.262]	[1.152]	[1.151]				0.364
Female	0.415	0.370	0.472	0.415	*0.785*	*0.620*	*0.533*	*0.845*
	[0.498]	[0.487]	[0.504]	[0.498]				*0.507*
Birth in Political Capital	0.359	0.352	0.283	0.302	*0.340*	*0.614*	*0.285*	*0.217*
	[0.484]	[0.482]	[0.455]	[0.464]				*0.336*
Asset Ownership	0.283	0.315	0.245	0.321	*0.882*	*0.659*	*0.234*	*0.524*
	[0.455]	[0.469]	[0.434]	[0.471]				*0.318*
Income	35273.774	40101.852	27849.057	33698.113	*0.781*	*0.156*	*0.068**	*0.198*
	[29089.252]	[30944.774]	[25649.559]	[24263.446]				*0.048***
Age	26.491	29.963	26.660	26.981	*0.203*	*0.321*	*0.722*	*0.575*
	[2.120]	[2.083]	[2.377]	[2.406]				*0.411*
Years of Education	14.793	15.148	15.038	15.321	*0.061 **	*0.396*	*0.568*	*0.425*
	[0.988]	[0.998]	[1.143]	[1.221]				*0.383*
Visited Foreign Country	0.208	0.222	0.245	0.226	*0.722*	*0.756*	*0.690*	*0.645*
	[0.409]	[0.420]	[0.434]	[0.423]				*0.956*
Occupational Group Designation								
Administrative Service Chiefs	0.226	0.074	0.208	0.170	*0.200*	*0.031** *	*0.390*	*0.795*
	[0.423]	[0.264]	[0.409]	[0.379]				*0.066**
Police Chiefs	0.132	0.111	0.057	0.094	*0.348*	*0.723*	*0.239*	*0.196*
	[0.342]	[0.317]	[0.233]	[0.295]				*0.348*
Federal Revenue Chiefs	0.189	0.259	0.226	0.208	*0.519*	*0.431*	*0.908*	*0.642*
	[0.395]	[0.442]	[0.423]	[0.409]				*0.685*
Foreign Service Chiefs	0.038	0.074	0.151	0.076	*0.159*	*0.751*	*0.045***	*0.037***
	[0.192]	[0.264]	[0.361]	[0.267]				*0.154*
All Other Occupational Groups	0.302	0.352	0.208	0.359	*0.953*	*0.391*	*0.076**	*0.293*
	[0.464]	[0.482]	[0.469]	[0.484]				*0.107*
Number of candidates (total=213)	53	54	53	53				

Notes: Individual averages. Standard deviations in brackets. p-values corresponding to F-statistics are presented in italics. *Significant at the 10 percent level, ** at the 5 percent level.

Table 2: Impact of Treatments on Altruism

	Altruism Game		*Charity Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.0652***	0.0602***	0.170*	0.203**
	(0.0237)	(0.0219)	(0.0887)	(0.0954)
Stand-alone Malleability (*M*)	-0.0204	-0.0220	-0.0185	-0.0229
	(0.0198)	(0.0192)	(0.0960)	(0.0969)
Joint Treatment (*UM*)	-0.00573	-0.0178	-0.0149	-0.0546
	(0.0102)	(0.0129)	(0.0959)	(0.0970)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
Mean of dep. var. (placebo)	0.498	0.498	0.604	0.604
p-value (test: *U = UM*)	0.004**	0.001**	0.035**	0.004**
p-value (test: *M = UM*)	0.485	0.849	0.967	0.716
p-value (test: *U = M*)	0.004**	0.002**	0.032**	0.012**
p-value (test: U*M = U + M*)	0.107	0.047**	0.180	0.056

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variables are normalized to an index between 0 and 1. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Field Outcomes - Blood Donations, Orphanage Visits and Volunteering

	Agreement to Donate		Appointment to Donate		Orphanage Visit	Volunteering in Schools
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.213**	-0.0335	0.261***	0.120	0.494***	0.236**
	(0.0990)	(0.124)	(0.0951)	(0.121)	(0.0942)	(0.103)
Stand-alone Malleability (*M*)	0.00707	0.00477	-0.0283	-0.0562	-0.00153	0.0332
	(0.0877)	(0.115)	(0.0832)	(0.110)	(0.0944)	(0.0970)
Joint Treatment (*UM*)	0.0880	0.0449	0.00195	-0.0575	0.0218	0.0590
	(0.0928)	(0.109)	(0.0842)	(0.0954)	(0.0935)	(0.0949)
Matching Blood Request (*T*)		-0.0703		-0.0297		
		(0.139)		(0.138)		
Matching Blood Request X Stand-alone Utilitarian (*U X T*)		0.530**		0.300		
		(0.207)		(0.206)		
Matching Blood Request X Stand-alone Malleability (*M X T*)		0.0365		0.0735		
		(0.188)		(0.190)		
Matching Blood Request X Joint Treatment (*UM X T*)		0.156		0.206		
		(0.220)		(0.215)		
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	207	207	207	207	213	213
Mean of dep. var. (placebo)	0.192	0.192	0.154	0.154	0.264	0.358
p-value (test: *U* = *UM*)	0.081	0.572	0.009**	0.302	0.087*	0.145
p-value (test: *M* = *UM*)	0.926	0.545	0.991	0.473	0.584	0.881
p-value (test: *U* = *M*)	0.058	0.994	0.008**	0.754	0.025**	0.185
p-value (test: U*M* = *U* + *M*)	0.058	0.922	0.020**	0.294	0.208	0.064*

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable in columns (1) and (2) are dummies that switch on for agreement to donate blood. The dependent variables in columns (3) and (4) are dummies for setting up an actual appointment for blood donation at a local blood bank. The dependent variable in columns (4) and (5) are dummies for choosing to visit orphanage and volunteering at impoverished schools relative to choice of attending a lecture by a senior bureaucrat. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Impact on Teamwork, Research Methods and Soft Skills Assessments - Standardized

	Teamwork Assessment		*Soft-Skills Assessment*		*Research Methods Assessment*	
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (U)	0.476**	0.479**	0.530**	0.547**	0.0587	0.115
	(0.189)	(0.201)	(0.223)	(0.249)	(0.199)	(0.210)
Stand-alone Malleability (M)	-0.0381	-0.0436	0.0555	0.0582	-0.101	-0.0811
	(0.196)	(0.214)	(0.178)	(0.181)	(0.198)	(0.205)
Joint Treatment (UM)	-0.0575	-0.0632	-0.164	-0.0784	0.0417	0.0809
	(0.195)	(0.205)	(0.145)	(0.157)	(0.194)	(0.197)
Individual Controls	No	Yes	No	Yes	No	Yes
Observations	213	213	213	213	213	213
p-value (test: $U = UM$)	0.007***	0.009***	0.038**	0.047**	0.420	0.348
p-value (test: $M = UM$)	0.920	0.925	0.149	0.364	0.459	0.410
p-value (test: $U = M$)	0.005***	0.006***	0.001**	0.004**	0.929	0.859
p-value (test: $UM = U + M$)	0.067**	0.083**	0.006**	0.012**	0.760	0.866

Note: Robust Newey-West standard errors appear in parenthesis. All dependent variables are standardized to mean 0 and standard deviation of 1. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. Dependent variables in Columns (1) and (2) present scores from regular public policy training courses on the original scale of 0 to 10 on the workshop *Teams & Group Decisions*. This workshop assesses policymakers team decisions. This assessment is marked by a committee of senior bureaucrats and academics. Dependent variables in Columns (3) and (4) present soft skills assessment on negotiations and leadership skills. Dependent variables in Columns (5) and (6) scores on *Quantitative Assessment (Research Methods)* are reported. This assessment content included a statistical inference course with emphasis on hypothesis testing, multivariate regression analysis with applications to policy-making, and randomized evaluations. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Impact of Treatments in Strategic Dilemmas

	Cooperation Game		*Coordination Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.140***	0.138***	0.0841**	0.0719*
	(0.0470)	(0.0504)	(0.0337)	(0.0365)
Stand-alone Malleability (M)	-0.0412	-0.0399	0.0278	0.0246
	(0.0403)	(0.0414)	(0.0299)	(0.0324)
Joint Treatment (*UM*)	-0.00251	-0.00907	0.0184	0.0155
	(0.0371)	(0.0410)	(0.0341)	(0.0346)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
Mean of dep. var. (placebo)	0.535	0.535	0.849	0.849
p-value (test: *U* = *UM*)	0.001**	0.002**	0.045**	0.088
p-value (test: *M* = *UM*)	0.264	0.405	0.748	0.803
p-value (test: *U* = *M*)	0.000**	0.000**	0.048**	0.093
p-value (test: U*M* = *U* + *M*)	0.087	0.083	0.048**	0.117

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is normalized to an index between 0 and 1 for cooperation and coordination respectively. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Alternative Mechanisms

	Competition Game	*Patience Game*	*Perseverance Game*	*Redistribution Game*	*Risk Aversion Game*	*Trust Game*
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian	0.124	-0.00139	-0.0662	0.00981	0.00174	0.495*
(*U*)	(0.0991)	(0.0184)	(0.0582)	(0.0112)	(0.0488)	(0.291)
Stand-alone Malleability	0.0258	-0.00887	-0.0547	0.0105	-0.0161	-0.163
(*M*)	(0.0982)	(0.0225)	(0.0616)	(0.00915)	(0.0540)	(0.287)
Joint Treatment (*UM*)	0.0600	-0.0136	0.0269	0.00835	-0.0514	-0.241
	(0.0990)	(0.0194)	(0.0724)	(0.00793)	(0.0556)	(0.265)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213	213
Mean of dep. var. (placebo)	0.321	0.604	0.132	0.492	0.732	0.538
p-value (test: *U* = *UM*)	0.658	0.462	0.165	0.434	0.270	0.822
p-value (test: *M* = *UM*)	0.662	0.804	0.210	0.780	0.499	0.236
p-value (test: *U* = *M*)	0.368	0.750	0.803	0.651	0.711	0.187
p-value (test: U*M* = *U* + *M*)	0.677	0.907	0.096	0.171	0.534	0.683

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is normalized to an index between 0 and 1 for behavioral games on competition, patience, perseverance, redistribution, risk and trust games. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix to:

Altruism in Governance:

Insights from Randomized Training for Pakistan's Junior Ministers

By Sultan Mehmood, Shaheen Naseer and Daniel Chen

Contents

Appendix A. Additional Robustness Checks

Appendix B. Experimental Setup, Scripts, Audio, and Additional Tables

Appendix A. Additional Robustness Checks Tables

Table A1: Impact of Treatments on Main Variables with Lasso Controls

	(1) *Altruism Game*	(2) *Charity Game*	(3) *Soft-Skills Assessment*	(4) *Teamwork Assessment*	(5) *Research Methods Assessment*
Stand-alone	0.0647***	0.170*	0.177**	0.694**	0.0862
Utilitarian	(0.0218)	(0.0885)	(0.0734)	(0.274)	(0.287)
(U)	[0.003]	[0.055]	[0.016]	[0.011]	[0.764]
Stand-alone	-0.0232	-0.0185	0.0185	-0.0556	-0.148
Malleability (*M*)	(0.0196)	(0.0956)	(0.0582)	(0.284)	(0.285)
	[0.237]	[0.846]	[0.750]	[0.845]	[0.603]
Joint Treatment	-0.0089	-0.0149	-0.0548	-0.0839	0.0611
(*UM*)	(0.0116)	(0.0956)	(0.0472)	(0.282)	(0.279)
	[0.443]	[0.876]	[0.245]	[0.766]	[0.827]
Controls (Lasso)	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213
R-squared	0.153	0.027	0.066	0.048	0.004
p-value (test: *U= UM*)	0.0015***	0.0364**	0.0006***	0.0041***	0.9299
p-value (test: *M = UM*)	0.4790	0.9696	0.1472	0.9199	0.4582
p-value (test: *U= M*)	0.0017***	0.0330**	0.0366**	0.0062***	0.4193
p-value (test: U*M = U + M*)	0.0911*	0.2019	0.0050***	0.0658*	0.7597

Note: Robust Newey-West standard errors appear in parenthesis, while the corresponding p-values are reported in square brackets. The dependent variables are normalized to an index between 0 and 1. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The control variables to be included in each regression are selected via the Post Double Selection Lasso approach. In column (1) the following Lasso selected control(s): foreign visits. In other columns no control is selected. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Impact of Treatments on Main Variables estimated via Double Debiased Machine Learning (DDML)

	(1) *Altruism Game*	(2) *Charity Game*	(3) *Soft-Skills Assessment*	(4) *Teamwork Assessment*	(5) *Research Methods Assessment*
Stand-alone	0.0615***	0.251***	0.110	0.508*	0.250
Utilitarian	(0.0177)	(0.0925)	(0.0743)	(0.2665)	(0.2727)
(U)	[0.001]	[0.007]	[0.138]	[0.057]	[0.359]
Stand-alone	-0.0275	0.0892	-0.0063	0.0976	-0.1404
Malleability (*M*)	(0.0188)	(0.0965)	(0.0561)	(0.2730)	(0.2829)
	[0.143]	[0.355]	[0.911]	[0.721]	[0.620]
Joint Treatment	-0.0018	0.0337	-0.0478	-0.0457	-0.1277
(*UM*)	(0.0077)	(0.0854)	(0.0546)	(0.2412)	(0.2465)
	[0.559]	[0.693]	[0.382]	[0.850]	[0.604]
Controls (Lasso)	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213
p-value (test: *U= UM*)	0.0012***	0.0226**	0.0261**	0.0401**	0.1762
p-value (test: *M = UM*)	0.2091	0.5882	0.4240	0.6136	0.9654
p-value (test: *U= M*)	0.0006***	0.1006	0.0980*	0.1363	0.1615
p-value (test: U*M = U + M*)	0.1489	0.0323**	0.0933*	0.1089	0.5750

Note: Robust Newey-West standard errors appear in parenthesis, while the corresponding p-values are reported in square brackets. The dependent variables are normalized to an index between 0 and 1. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The DDML estimation assumes a partially linear model and implements the cross-fitting algorithm. All 14 control variables are included to estimate the orthogonalized version of the outcome variables and treatment variables of interest.

Table A3: Results from the Field - Blood Donations, Orphanage Visits and Volunteering with Lasso Controls

	(1)	(2)	(3)	(4)	(5)	(6)
	Agreement to Donate		Appointment to Donate		Orphanage Visit	Volunteering in Schools
Stand-alone Utilitarian (*U*)	0.209**	-0.0641	0.248***	0.0769	0.491***	0.241***
	(0.0946)	(0.110)	(0.0938)	(0.119)	(0.0844)	(0.0924)
	[0.027]	[0.559]	[0.008]	[0.517]	[0.000]	[0.009]
Stand-alone Malleability (*M*)	0.0145	-0.0150	-0.0229	-0.0584	0	0.0370
	(0.0865)	(0.111)	(0.0818)	(0.103)	(0.0862)	(0.0878)
	[0.867]	[0.893]	[0.780]	[0.572]	[1.000]	[0.673]
Joint Treatment (*UM*)	0.0769	0.0214	-0.0192	-0.0919	0.0241	0.0618
	(0.0900)	(0.108)	(0.0824)	(0.0949)	(0.0877)	(0.0892)
	[0.393]	[0.893]	[0.815]	[0.333]	[0.783]	[0.488]
Matching Blood Request		-0.103		-0.0769		
		(0.128)		(0.126)		
		[0.421]		[0.543]		
Matching Blood Request X Stand-alone Utilitarian (*U X T*)		0.619***		0.394**		
		(0.177)		(0.186)		
		[0.000]		[0.034]		
Matching Blood Request X Stand-alone Malleability (*M X T*)		0.111		0.113		
		(0.176)		(0.168)		
		[0.527]		[0.502]		
Matching Blood Request X Joint Treatment (*UM X T*)		0.200		0.251		
		(0.193)		(0.183)		
		[0.301]		[0.170]		
Controls (Lasso)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	207	207	207	207	213	213
R-squared	0.032	0.112	0.092	0.127	0.180	0.037
p-value (test: *U = UM*)	0.1678	0.4355	0.0030***	0.1230	0.0000***	0.0612*
p-value (test: *M = UM*)	0.4781	0.7434	0.9626	0.7173	0.7854	0.7864
p-value (test: *U = M*)	0.0356**	0.6603	0.0024***	0.2468	0.0000***	0.0308**
p-value (test: U*M = U + M*)	0.2558	0.5202	0.0441**	0.4629	0.0001***	0.0958*

Note: Robust Newey-West standard errors appear in parenthesis, while corresponding p-values are reported in square brackets. The dependent variables in columns (1) and (2) are dummies that switch on for agreement to donate blood. The dependent variables in columns (3) and (4) are dummies for setting up an actual appointment for blood donation at a local blood bank. The dependent variables in columns (4) and (5) are dummies for choosing to visit orphanage and volunteering at impoverished schools relative to choice of attending a lecture by a senior bureaucrat. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The control variables to be included in each regression are selected via the Post Double Selection Lasso approach. In all columns no control is selected. *** p<0.01, ** p<0.05, * p<0.1

Table A4: Results from the Field - Blood Donations, Orphanage Visits and Volunteering via Double Debiased Machine Learning (DDML)

	(1)	(2)	(3)	(4)	(5)	(6)
	Agreement to Donate		Appointment to Donate		Orphanage Visit	Volunteering in Schools
Stand-alone Utilitarian (*U*)	0.1638*	-0.0366	0.2729***	0.1116	0.3756***	0.1269bo
	(0.0904)	(0.1138)	(0.0929)	(0.1129)	(0.0870)	(0.1000)
	[0.070]	[0.748]	[0.003]	[0.323]	[0.000]	[0.204]
Stand-alone Malleability (*M*)	0.0382	0.1085	0.0085	0.0523	-0.0909	-0.0379
	(0.0887)	(0.1081)	(0.0829)	(0.100)	(0.0856)	(0.0924)
	[0.667]	[0.315]	[0.918]	[0.602]	[0.289]	[0.682]
Joint Treatment (*UM*)	0.1002	0.0547	0.0247	-0.0008	-0.0265	-0.0340
	(0.0819)	(0.0880)	(0.0771)	(0.0829)	(0.0857)	(0.0882)
	[0.221]	[0.534]	[0.749]	[0.992]	[0.757]	[0.700]
Matching Blood Request		0.0352		0.0954		
		(0.1146)		(0.1117)		
		[0.758]		[0.393]		
Matching Blood Request X Stand-alone Utilitarian (*U X T*)		0.3611**		0.2602		
		(0.1781)		(0.1780)		
		[0.043]		[0.144]		
Matching Blood Request X Stand-alone Malleability (*M X T*)		-0.2052		-0.1587		
		(0.1515)		(0.1448)		
		[0.175]		[0.273]		
Matching Blood Request X Joint Treatment (*UM X T*)		0.1912		0.1138		
		(0.1696)		(0.1767)		
		[0.259]		[0.520]		
Controls (Lasso)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	207	207	207	207	213	213
p-value (test: *U* = *UM*)	0.5149	0.4320	0.0103**	0.3801	0.0000***	0.1066
p-value (test: *M* = *UM*)	0.5078	0.6201	0.8516	0.5838	0.4912	0.9681
p-value (test: *U* = *M*)	0.1622	0.2164	0.0029***	0.6041	0.0000***	0.0988*
p-value (test: U*M* = *U* + *M*)	0.4667	0.9183	0.0589*	0.2918	0.0188**	0.3861

Note: Robust Newey-West standard errors appear in parenthesis, while corresponding p-values are reported in square brackets. The dependent variables in columns (1) and (2) are dummies that switch on for agreement to donate blood. The dependent variables in columns (3) and (4) are dummies for setting up an actual appointment for blood donation at a local blood bank. The dependent variables in columns (4) and (5) are dummies for choosing to visit orphanage and volunteering at impoverished schools relative to choice of attending a lecture by a senior bureaucrat. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The DDML estimation assumes a partially linear model and implements the cross-fitting algorithm. All 14 control variables are included to estimate the orthogonalized version of the outcome variables and treatment variables of interest. *** p<0.01, ** p<0.05, * p<0.1

Table A5: Balance test by Book Assignment - Before Book Assignment

	(1) *Altruism Game*	(2) *Charity Game*	(3) *Agreement to Donate*	(4) *Appointment to Donate*
U	0.0917	0.305	0.178	0.0366
	(0.0606)	(0.204)	(0.237)	(0.265)
	[0.132]	[0.136]	[0.454]	[0.890]
M	-0.0372	-0.150	0.170	-0.0593
	(0.0476)	(0.201)	(0.179)	(0.183)
	[0.436]	[0.456]	[0.343]	[0.747]
UM	0.00969	0.0375	0.0272	-0.109
	(0.0218)	(0.176)	(0.157)	(0.160)
	[0.657]	[0.831]	[0.862]	[0.497]
Book Assigned	0.0243	0.100	0.172	-0.0413
	(0.0216)	(0.205)	(0.187)	(0.186)
	[0.262]	[0.625]	[0.357]	[0.824]
U X Book Assigned	-0.0469	0.189	0.0500	0.334
	(0.0498)	(0.291)	(0.259)	(0.269)
	[0.347]	[0.517]	[0.847]	[0.216]
M X Book Assigned	-0.00294	-0.0735	-0.373	-0.0646
	(0.0477)	(0.290)	(0.276)	(0.276)
	[0.951]	[0.800]	[0.178]	[0.815]
UM X Book Assigned	-0.0833**	0.0889	-0.212	-0.133
	(0.0360)	(0.264)	(0.256)	(0.266)
	[0.022]	[0.737]	[0.408]	[0.617]
Controls	Yes	Yes	Yes	Yes
Observations	213	213	207	207
R-squared	0.269	0.195	0.133	0.148

Note: Robust Newey-West standard errors appear in parenthesis, while corresponding p-values are reported in square brackets. The dependent variable in column (3) is a dummy that switches on for agreement to donate blood. The dependent variable in column (4) is a dummy for setting up an actual appointment for blood donation at a local blood bank. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. Book assigned is a dummy variable that switches when the empathy book is assigned. The estimations obtained from OLS regressions includes the following controls: empathy book chosen, interaction of empathy book chosen with all the treatments, written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Heterogeneous Impact of Treatments Evaluations, Orphanage Visits and Volunteering by Book Assignment - After Book Assignment

	(1) *Soft-Skills Assessment*	(2) *Teamwork Assessment*	(3) *Research Methods Assessment*	(4) *Orphanage Visit*	(5) *Volunteering in Schools*
U	0.281	0.649	-0.103	0.155	-0.163
	(0.226)	(0.716)	(0.789)	(0.224)	(0.210)
	[0.216]	[0.314]	[0.885]	[0.447]	[0.391]
M	0.122	0.152	-0.719	-0.00852	0.0617
	(0.154)	(0.689)	(0.600)	(0.178)	(0.194)
	[0.428]	[0.811]	[0.197]	[0.959]	[0.733]
UM	-0.0778	0.934	-0.187	-0.0374	0.00492
	(0.0996)	(0.608)	(0.549)	(0.170)	(0.174)
	[0.436]	[0.103]	[0.715]	[0.813]	[0.976]
Book	-0.0103	0.683	-0.777	-0.00310	-0.0204
Assigned	(0.0858)	(0.498)	(0.535)	(0.164)	(0.167)
	[0.904]	[0.137]	[0.114]	[0.984]	[0.894]
U X Book	-0.0263	-0.365	-0.000271	0.0313	0.394
Assigned	(0.266)	(0.977)	(0.856)	(0.256)	(0.286)
	[0.922]	[0.667]	[0.999]	[0.891]	[0.118]
M X Book	-0.231	-0.948	1.426*	-0.00745	-0.0346
Assigned	(0.160)	(0.786)	(0.825)	(0.254)	(0.258)
	[0.151]	[0.184]	[0.058]	[0.974]	[0.883]
UM X Book	0.140	-0.793	0.643	0.284	0.295
Assigned	(0.117)	(0.682)	(0.684)	(0.225)	(0.224)
	[0.232]	[0.206]	[0.307]	[0.168]	[0.149]
Controls	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213
R-squared	0.173	0.133	0.103	0.269	0.139

Note: Robust Newey-West standard errors appear in parenthesis, while corresponding p-values are reported in square brackets. The dependent variables in columns (4) and (5) are dummies for choosing to visit orphanage and volunteering at impoverished schools relative to choice of attending a lecture by a senior bureaucrat. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. Book assigned is a dummy variable that switches when the empathy book is assigned. The estimations obtained from OLS regressions includes the following controls: empathy book chosen, interaction of empathy book chosen with all the treatments, written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Impact of Treatments on main variables standardized with respect to placebo group

	(1) *Altruism Game*	(2) *Charity Game*	(3) *Soft-Skills Assessment*	(4) *Teamwork Assessment*	(5) *Research Methods Assessment*
Stand-alone Utilitarian (*U*)	1.392***	0.412**	0.624**	0.474**	0.114
	(0.481)	(0.188)	(0.268)	(0.191)	(0.197)
	[0.004]	[0.029]	[0.021]	[0.014]	[0.564]
Stand-alone Malleability (*M*)	-0.509	-0.0465	0.0664	-0.0431	-0.0798
	(0.429)	(0.191)	(0.192)	(0.204)	(0.192)
	[0.237]	[0.808]	[0.730]	[0.833]	[0.678]
Joint Treatment (*UM*)	-0.411	-0.111	-0.0894	-0.0625	0.0797
	(0.287)	(0.190)	(0.167)	(0.195)	(0.185)
	[0.154]	[0.559]	[0.593]	[0.749]	[0.667]
Controls	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213
R-squared	0.250	0.144	0.137	0.084	0.073
p-value (test: *U* = *UM*)	0.0001	0.0047	0.0036	0.0837	0.2336
p-value (test: *M* = *UM*)	0.8260	0.7302	0.3635	0.7596	0.9879
p-value (test: *U* = *M*)	0.0023	0.0132	0.0467	0.1040	0.2079
p-value (test: U*M* = *U*+*M*)	0.0523	0.0742	0.0105	0.0819	0.1539

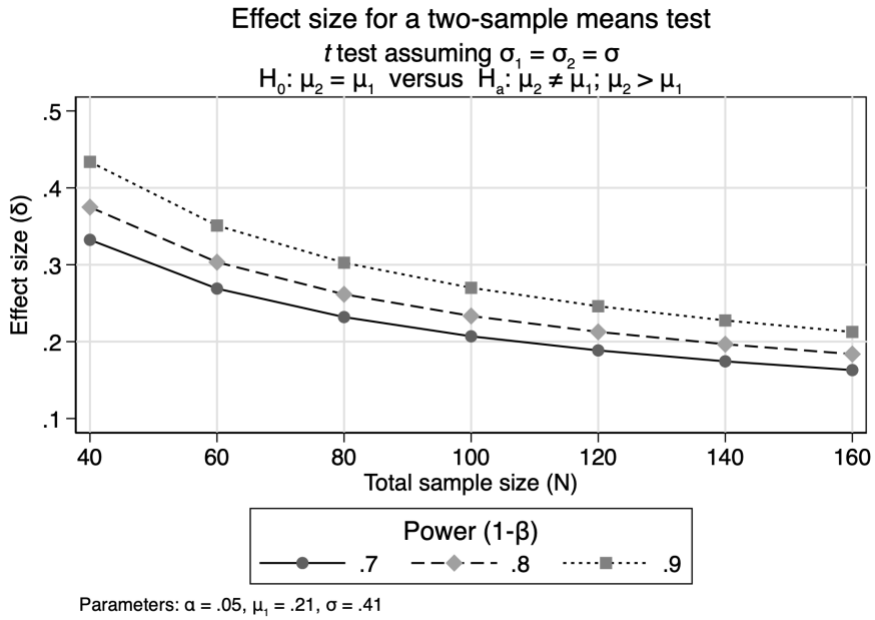
Note: Robust Newey-West standard errors appear in parenthesis, while corresponding p-values are reported in square brackets. The dependent variables are normalized to an index between 0 and 1 with respect to the placebo group. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following Lasso selected control(s): foreign visits. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Results from the Field - Blood Donations, Orphanage Visits and Volunteering standardized with respect to placebo group

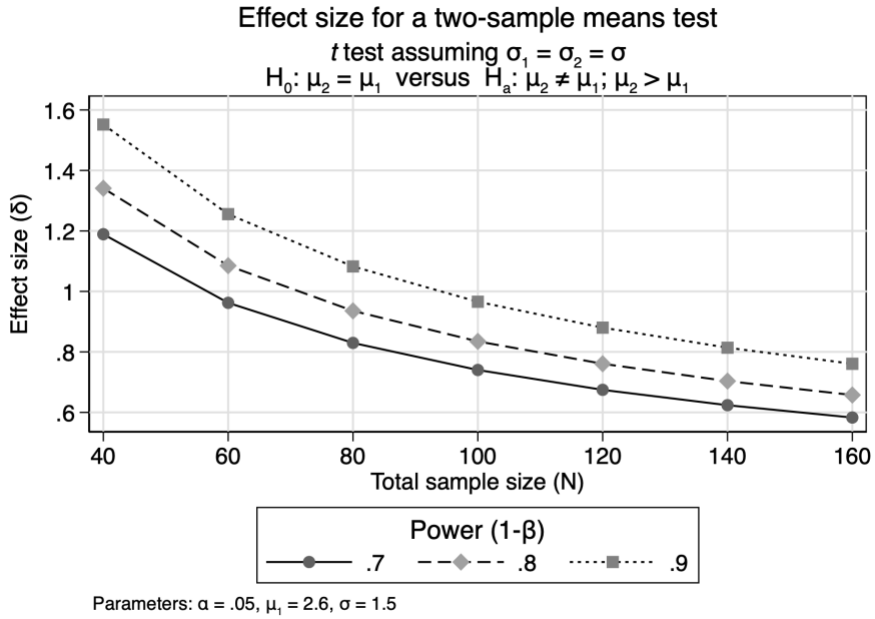
	(1)	(2)	(3)	(4)	(5)	(6)
	Agreement to Donate		Appointment to Donate		Orphanage Visit	Volunteering in Schools
Stand-alone Utilitarian	0.500**	-0.0787	0.633***	0.290	1.094***	0.522**
(*U*)	(0.226)	(0.283)	(0.224)	(0.285)	(0.197)	(0.216)
	[0.028]	[0.781]	[0.005]	[0.309]	[0.001]	[0.017]
Stand-alone Malleability	0.0166	0.0112	-0.0685	-0.136	-0.00337	0.0734
(*M*)	(0.202)	(0.265)	(0.198)	(0.260)	(0.198)	(0.205)
	[0.935]	[0.966]	[0.729]	[0.600]	[0.986]	[0.720]
Joint Treatment (*UM*)	0.207	0.105	0.00474	-0.140	0.0483	0.131
	(0.211)	(0.249)	(0.197)	(0.224)	(0.197)	(0.200)
	[0.328]	[0.672]	[0.981]	[0.535]	[0.807]	[0.516]
Matching Blood Request		-0.165		-0.0721		
		(0.309)		(0.314)		
		[0.594]		[0.819]		
Matching Blood Request		1.246***		0.728		
X Stand-alone Utilitarian		(0.459)		(0.470)		
(*U X T*)		[0.007]		[0.123]		
Matching Blood Request		0.0857		0.178		
X Malleability (*M X T*)		(0.418)		(0.435)		
		[0.838]		[0.682]		
Matching Blood Request		0.367		0.499		
X Joint Treatment		(0.485)		(0.486)		
(*UM X T*)		[0.449]		[0.302]		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	207	207	207	207	213	213
R-squared	0.103	0.163	0.118	0.148	0.226	0.093
p-value (test: *U* = *UM*)	0.2112	0.5083	0.0056	0.1173	0.0001	0.0715
p-value (test: *M* = *UM*)	0.3736	0.7215	0.7117	0.9890	0.7929	0.7808
p-value (test: *U* = *M*)	0.0286	0.7570	0.0017	0.1497	0.0001	0.0394
p-value (test: U*M* = *U* + *M*)	0.3274	0.6569	0.0647	0.4384	0.0002	0.1231

Note: Robust Newey-West standard errors appear in parenthesis, corresponding while p-values are reported in square brackets. The dependent variables in columns (1) and (2) are dummies that switch on for agreement to donate blood. The dependent variables in columns (3) and (4) are dummies for setting up an actual appointment for blood donation at a local blood bank. The dependent variables in columns (4) and (5) are dummies for choosing to visit orphanage and volunteering at impoverished schools relative to choice of attending a lecture by a senior bureaucrat. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: interview test scores and birth in political capitals. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Power Calculation Graphs
Panel A; Appointment to Donate



Panel B: Teamwork Assessments



Note: The figure above provides the average effect size for Appointment to Donate in Panel A and for Teamwork Assessments in Panel B. These two outcomes are chosen since they provide the largest deviation in the MDEs computed.

Appendix B. Experimental Setup, Scripts, and Additional Tables

Table B2: Script of Email sent

Subject: Workshop - Material

Full Video, Audio along with transcripts of each treatment is also available below:

Utilitarian Treatment Full Audio, Video and Transcript (17 mins, 53 seconds):
Malleability Treatment Full Audio, Video and Transcript (12 mins, 49 seconds):
Combined Treatment Full Audio, Video and Transcript (17 mins, 53 seconds):

Table B6: Script of the Structured Discussion Post-Lectures

Candidate 1:

Q1. What do you think were the main messages of the lecture? Q2. How do you think you may apply lessons from today's lecture in your career? Give at least 3 examples.

Candidate 2:

Q1. What struck you most about today's lectures and why? Please be specific on what you think are the key takeaways of today's lectures. Q2. Can you give three examples on how the lessons of today's workshop could be applied in your official duties?

Table B7: Attrition in Blood Donation Responses

	Drop-Outs (not answering calls for blood donations)		
	(1)	(2)	(3)
Stand-alone Utilitarian (*U*)	0.0302		0.0465
	(0.0474)		(0.0700)
Stand-alone Malleability (*M*)	-0.00858		0.00132
	(0.0346)		(0.0523)
Joint Treatment (*UM*)	0.0102		0.0197
	(0.0410)		(0.0544)
Matching Blood Request			-0.0422
			(0.0470)
Matching Blood Request X Stand-alone Utilitarian (*U X T*)		-0.0181	-0.0121
		(0.0436)	(0.0781)
Matching Blood Request X Stand-alone Malleability (*M X T*)		-0.0531**	-0.00234
		(0.0240)	(0.0628)
Matching Blood Request X Joint Treatment (*UM X T*)		-0.0540*	-0.0232
		(0.0287)	(0.0570)
Individual Controls	Yes	Yes	Yes
Observations	213	213	213
Mean of dep. var. (placebo)	0.083	0.089	0.099
p-value (test: *U* = *UM*)	0.635	0.409	0.693
p-value (test: *M* = *UM*)	0.496	0.971	0.698
p-value (test: *U* = *M*)	0.369	0.419	0.521
p-value (test: U*M* = *U* + *M*)	0.818	0.720	0.729

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is a dummy for not answering phone calls for blood donation. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies.*** p<0.01, ** p<0.05, * p<0.1.

Table B8: Impact on Standardized Outcome Variables

	Altruism Game		*Charity Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.560*** (0.203)	0.517*** (0.188)	0.364* (0.186)	0.435** (0.200)
Stand-alone Malleability (*M*)	-0.175 (0.170)	-0.189 (0.165)	-0.0319 (0.201)	-0.0410 (0.203)
Joint Treatment (*UM*)	-0.0492 (0.0875)	-0.153 (0.111)	-0.0234 (0.201)	-0.108 (0.203)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
Mean of dep. var. (placebo)	-0.064	-0.064	-0.083	-0.083
p-value (test: *U* = *UM*)	0.004***	0.002***	0.039**	0.007***
p-value (test: *M* = *UM*)	0.478	0.833	0.967	0.739
p-value (test: *U* = *M*)	0.004***	0.004***	0.035**	0.017**
p-value (test: U*M* = *U* + *M*)	0.108	0.064*	0.196	0.078*

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is standardized to have a mean of zero and a standard deviation of one. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table B9: Mechanism - Impact of Treatments on Decision Making - Standardized

	Cooperation Game		*Coordination Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (U)	0.636***	0.624***	0.514**	0.439*
	(0.213)	(0.229)	(0.206)	(0.223)
Stand-alone	-0.187	-0.181	0.170	0.150
Malleability (*M*)	(0.183)	(0.188)	(0.182)	(0.198)
Joint Treatment (*UM*)	-0.0114	-0.0411	0.112	0.0948
	(0.168)	(0.186)	(0.209)	(0.211)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
Mean of dep. var. (placebo)	-0.185	-0.185	-0.172	-0.172
p-value (test: *U* = *UM*)	0.001***	0.003***	0.050*	0.096*
p-value (test: *M* = *UM*)	0.269	0.423	0.750	0.770
p-value (test: *U* = *M*)	0.000***	0.000***	0.054*	0.118
p-value (test: U*M* = *U* + *M*)	0.085*	0.093*	0.038**	0.099*

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is standardized to have a mean of zero and a standard deviation of one. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint Treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table B10: Exploratory Analysis – Alternative Mechanisms – Standardized

	Competitiveness Game	*Patience Game*	*Perseverance Game*	*Redistribution Game*	*Risk Aversion Game*	*Trust Game*
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.256	-0.0143	-0.217	0.226	0.00675	0.359*
	(0.205)	(0.190)	(0.191)	(0.258)	(0.189)	(0.211)
Stand-alone Malleability (*M*)	0.0534	-0.0914	-0.179	0.241	-0.0626	-0.118
	(0.203)	(0.232)	(0.202)	(0.211)	(0.210)	(0.208)
Joint Treatment (*UM*)	0.124	-0.140	0.0883	0.192	-0.200	-0.175
	(0.204)	(0.200)	(0.238)	(0.182)	(0.216)	(0.192)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	213	213
Mean of dep. var. (placebo)	-0.107	0.187	0.090	-0.197	0.011	0.063
p-value (test: $U = UM$)	0.658	0.462	0.165	0.434	0.270	0.822
p-value (test: $M = UM$)	0.662	0.804	0.210	0.780	0.499	0.236
p-value (test: $U = M$)	0.368	0.750	0.803	0.651	0.711	0.187
p-value (test: $UM = U + M$)	0.677	0.907	0.096	0.171	0.534	0.683

Note: Robust Newey-West standard errors appear in parenthesis. The dependent variable is standardized to have a mean of zero and a standard deviation of one. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table B11: Effect of Treatments on the Importance of Prosociality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Risk Tolerance	*Patience*	*Perseverance*	*Altruism*	*Trust in others*	*Preference for redistribution*	*Cooperation*	*Competition*
Utilitarian (U)	-0.120	-0.0200	-0.0400	0.0600	-0.120	0.0400	-0.0645	-0.560**
	(0.235)	(0.0721)	(0.0992)	(0.121)	(0.234)	(0.112)	(0.0798)	(0.228)
Malleability	-0.126	-0.0487	-0.0862	-0.0506	-0.238	-0.0917	-0.104	-0.338
(M)	(0.224)	(0.0735)	(0.107)	(0.129)	(0.222)	(0.128)	(0.0849)	(0.231)
Joint	-0.304	-0.0331	0.01000	0.132	-0.0723	-0.0122	-0.0361	-0.937***
Treatment (UM)	(0.225)	(0.0725)	(0.0961)	(0.112)	(0.225)	(0.120)	(0.0817)	(0.206)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	205	205	204	205	205	204	203	203

Note: Robust Newey-West standard errors appear in parenthesis. Dependent variables in Columns 1-8 are a rating on a scale of 1 to 4 with 1 being "not important at all" and 4 being "very important" on different traits with the statement "*How important do you think the following traits are? Risk tolerance, patience, perseverance, altruism, trust in others, preference for redistribution, cooperation and competition.*" U, M and UM are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimates are the OLS regressions with the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. *** p<0.01, ** p<0.05, * p<0

Table B12: Average Effect Sizes estimates - Summing Up

	Altruism	*Perspective Taking*	*Field Measures*	*Policy Assessments*
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian	0.914***	0.519***	0.534***	0.377***
	(0.235)	(0.152)	(0.120)	(0.125)
Stand-alone Malleability	-0.256	-0.022	0.095	-0.025
	(0.234)	(0.140)	(0.115)	(0.108)
Joint Treatment	-0.244	0.023	0.117	-0.068
	(0.169)	(0.142)	(0.115)	(0.098)
Observations	213	213	207	213

Note: All estimates are average effect size estimates. In Column (1), Altruism is based on normalized dictator and charity games. Column (2) summarizes Perspective Taking which is based on coordination and cooperation games. Column (3) compiles our Field Measures which arebased on dummies for blood donations, for setting up an appointment to donate blood, orphanage field visit and volunteering in impoverished schools. Column (4) contains the average effect of Policy Assessments that is based on soft skills, teamwork, and research methods assessments. U, M and UM are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations are average standardized effect size using the seemingly unrelated regression framework to account for covariance across estimates. The following controls are: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits, and occupational group dummies. Standard errors in parentheses are clustered at individual level. Robust standard errors appear in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0$.

Table B13: Randomization Inference – With right-tailed p-values

	(1)	(2)	(4)	(5)	(6)	(7)
	Altruism Game	*Charity Game*	*Soft-Skills Assessment*	*Agreement to Donate*	*Appointment to Donate*	*Orphanage Visit*
Stand-alone Utilitarian (*U*)	0.060	0.223	0.183	0.213	0.261	0.494
	(0.004) ***	(0.015) **	(0.021) **	(0.028) **	(0.005) ***	(0.001) ***
	{0.010} ***	{0.010} ***	{0.002} ***	{0.015} ***	{0.004} ***	{0.000} ***
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	207	207	213
Mean of dep. var. (placebo)	0.498	0.604	0.509	0.216	0.176	0.264

Note: p-values corresponding to clustered standard errors at individual level appear in parenthesis, while the right-tailed p-values from permutation inference are reported in curly brackets. U is a dummy variable indicating randomly assigned Utilitarian treatment. All estimations include the following controls: written test scores, interview test scores, gender, birth in political capitals, asset ownership, income before joining civil service, age, education, foreign visits and occupational group dummies. M and UM i.e. Malleability and Joint treatment lectures are also added as controls as in the baseline specification. *ritest* in Stata is implemented with 1000 iterations to perform the permutation inference test. *** p<0.01, ** p<0.05, * p<0.1.

Table B14: Adjusting Multiple Hypothesis Testing

	Altruism Game	Charity Game	Cooperation Game	Coordination Game	Competition Game	Patience Game	Perseverance Game	Redistribution Game	Risk Aversion Game	Trust Game	Appointment to donate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Stand-alone Utilitarian (*U*)	0.060	0.223	0.138	0.072	0.124	-0.001	-0.066	0.010	0.002	0.495	0.261
p-value	(0.004)***	(0.015)**	(0.005)***	(0.040)**	(0.189)	(0.937)	(0.238)	(0.370)	(0.971)	(0.076)*	(0.005)***
Sharpened q-value	[0.060]*	[0.126]	[0.060]*	[0.300]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.553]	[0.060]*
FWER p-value	{0.144}	{0.225}	{0.144}	{0.366}	{0.902}	{1.000}	{0.941}	{0.990}	{1.000}	{0.562}	{0.145}
Stand-alone Malleability (*M*)	-0.022	-0.004	-0.040	0.025	0.026	-0.009	-0.055	0.010	-0.016	-0.163	-0.028
p-value	(0.237)	(0.962)	(0.317)	(0.419)	(0.782)	(0.684)	(0.358)	(0.237)	(0.758)	(0.554)	(0.729)
Sharpened q-value	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
FWER p-value	{0.941}	{1.000}	{0.981}	{0.995}	{1.000}	{1.000}	{0.987}	{0.941}	{1.000}	{0.999}	{1.000}
Joint Treatment (*UM*)	-0.018	-0.040	-0.009	0.016	0.060	-0.014	0.027**	0.008***	-0.051***	-0.241***	0.002***
p-value	(0.154)	(0.666)	(0.818)	(0.638)	(0.525)	(0.468)	(0.698)	(0.279)	(0.336)	(0.344)	(0.981)
Sharpened q-value	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
FWER p-value	{0.835}	{1.000}	{1.000}	{1.000}	{0.998}	{0.998}	{1.000}	{0.965}	{0.987}	{0.987}	{1.000}
Sample Size	213	213	213	213	213	213	213	213	213	213	213

Note: The baseline p-values corresponding to robust standard errors clustered at individual level appear in parenthesis, Anderson's sharpened q-values appear in square brackets, and List et al. (2019) FWER adjusted p-values appear in curly brackets. The adjusted p-values are computed under the most strident criteria possible i.e. nesting all 36 outcomes in a single family. The dependent variables for all games are normalized to an index between 0 and 1. U, M and UM are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. *** p<0.01, ** p<0.05, * p<0.1

Table B15: Robustness Check – Robustness Check – Design Analysis

		True Effect Size / Estimated Effect Size				
		(1) 100%	(2) 75%	(3) 50%	(4) 25%	(5) 12.5%
Table 2 Estimates of U Effect						
Col 2: 0.0602*** (0.0219)	Type S Error	0.000001	0.0000535	0.0015291	0.0384268	0.16695
	Type M Error	1.130	1.368066	1.877742	3.518828	6.918589
Col 4: 0.203** (0.0954)	Type S Error	0.0000384	0.000526	0.0066947	0.0765364	0.2237361
	Type M Error	1.324934	1.638602	2.343	4.498049	8.860003
Table 3 Estimates of U Effect						
Col 1: 0.213** (0.0990)	Type S Error	0.0000341	0.0004827	0.0063317	0.0745994	0.2213441
	Type M Error	1.31429	1.625218	2.318368	4.439929	8.706265
Col 3: 0.261*** (0.0951)	Type S Error	0.000001	0.0000544	0.0015453	0.0386176	0.1673068
	Type M Error	1.135224	1.348737	1.867235	3.546227	6.844305
Col 5: 0.494*** (0.0942)	Type S Error	2.92e-13	1.94e-09	3.09e-06	0.002071	0.0443306
	Type M Error	1.002201	1.013124	1.160564	1.941738	3.689501
Col 6: 0.236** (0.103)	Type S Error	0.0000169	0.00029	0.00455	0.0640373	0.2075745
	Type M Error	1.262471	1.55429	2.192688	4.17772	8.227695
Table 4 Estimates of U Effect						
Col 2: 0.476** (0.189)	Type S Error	0.0000106	0.0002071	0.0036595	0.057871	0.1988761
	Type M Error	1.234378	1.511277	2.123128	4.061346	7.944997
Col 4: 0.0602** (0.0219)	Type S Error	0.0000272	0.0004094	0.0056897	0.0710139	0.2168132
	Type M Error	1.293593	1.6033	2.294495	4.389435	8.537121
Col 6: 0.115 (0.210)	Type S Error	0.071504	0.1276664	0.2174408	0.3452093	0.4206604
	Type M Error	4.390875	5.756453	8.662965	17.14997	34.3918
Table 5 Estimates of U Effect						
Col 2: 0.138** (0.0504)	Type S Error	1.68e-06	0.0000557	0.0015691	0.0388973	0.1678281
	Type M Error	1.14063	1.356588	1.880845	3.582784	6.900623
Col 4: 0.0719* (0.0365)	Type S Error	0.0000843	0.0009324	0.0097082	0.0907083	0.2401985
	Type M Error	1.401949	1.771854	2.538932	4.885497	9.527714

Note: For each estimated effect size and standard error of the Stand-alone Utilitarian (U) training, we *estimate* further the probability of a sign error (Type S Error) and the potential exaggeration ratio in effect size (Type M Error) following the procedure proposed by Gelman and Carlin (2014), by considering the true effect size to be 100%, 75%, 50%, 25% and 12.5% of the estimates in our paper.

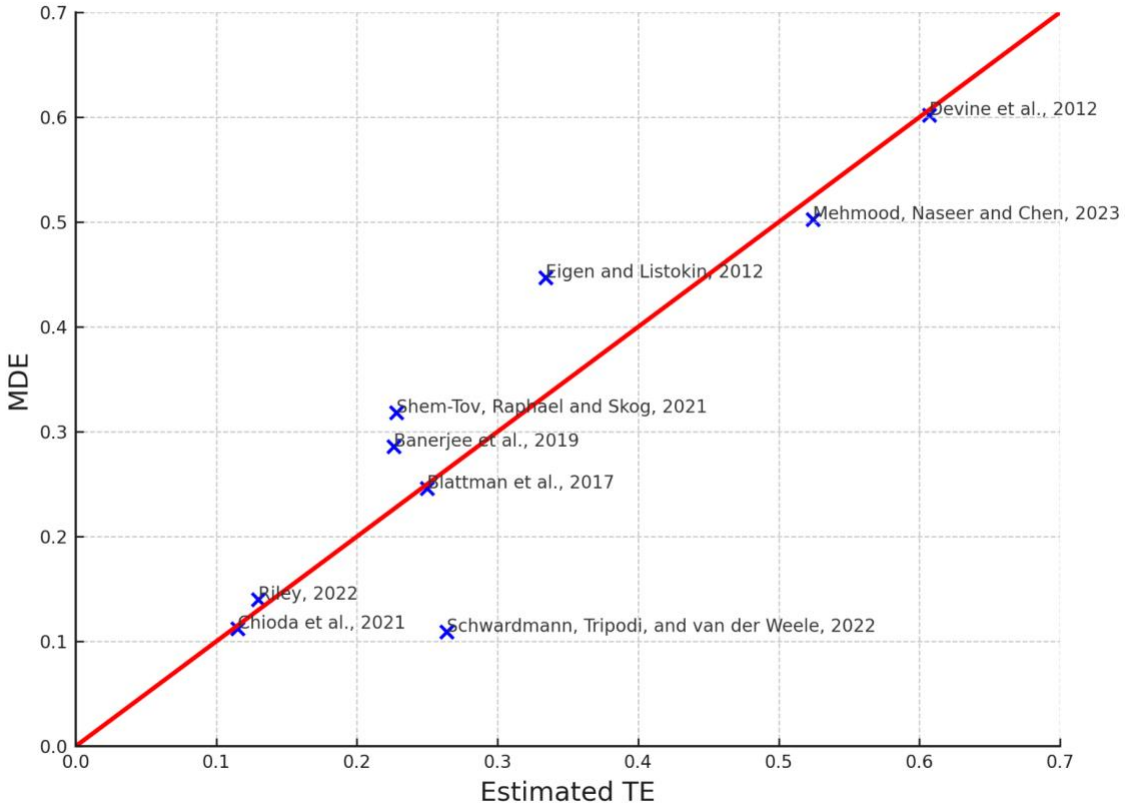
Table B16: Minimum Detectable Effects (MDE) across Relevant Studies

Paper	Intervention	Sample Size	Main Results	MDE	TE
Devine et al., 2012	A multi-faceted prejudice habit-breaking intervention.	91 non-Black introductory psychology students (67% female, 85% White), with 53 in treatment group and 38 in control.	Following the manipulation, treated participants had 0.19 lower IAT scores (equal to -0.607 SD, with std SE 0.215) than control group participants.	0.215 * 2.8 = 0.602	-0.607 SD
Riley, 2022	Female role-model building intervention involving cinema screening of the movie *Queen of Katwe*.	In the S4 class, 391 treated and 342 controlled. In the S6 class, 370 treated and 341 controlled.	For upper secondary school students, treatment 1 month before their exams results in an increase in their total exam score of 0.13 (se 0.05) standard deviations.	0.05 * 2.8 = 0.14	0.13 SD
Banerjee et al., 2019	Edutainment treatment screening TV series MTV *Shuga*.	54 screening centers that showed Shuga (treatment) and 26 that showed a "placebo" TV series. Among the attendees 63 people per center were randomly selected.	*Shuga* intervention reduced men's positive attitude towards gender-based violence (GBV) by 0.226 SD (se 0.102).	0.102 * 2.8 = 0.286	-0.226 SD
Eigen and Listokin, 2012	Randomly assign law school students to the role of petitioner or respondent in moot court competitions.	77 participants were assigned to respondent role and 96 to petitioner.	Being randomly assigned to the role of petitioner is associated with a -0.3343 SD (se 0.1597) decrease in the merits-based and moral confidence differential.	0.1597 * 2.8 = 0.4472	-0.3343 SD
Schwardmann, Tripodi, and van der Weele, 2022	Randomly assign experienced and motivated debaters to argue one side of a topical motion at international debating competitions.	473 debaters.	0.264 SD (se 0.039) gap in pre-debate factual beliefs between proposition and opposition debaters.	0.039 * 2.8 = 0.1092	0.264 SD
Shem-Tov, Raphael and Skog, 2021	Eligible youths were randomly assigned to participate in the Make-it-Right (MIR) restorative justice program or a control group where they faced standard criminal prosecution.	143 youth, 99 were assigned to MIR, and 44 faced regular felony prosecution.	After 1 year, the likelihood of rearresting of the MIR participants decreased by 0.228 (se 0.111), and after 4 years by 0.363 (se 0.165).	0.111 * 2.8 = 0.318 0.165 * 2.8 = 0.462	0.228 SD 0.363 SD
Blattman et al., 2017	A combination of Cognitive Behavioral Therapy (CBT) and unconditional cash transfers. The CBT aimed to reduce self-destructive beliefs or behaviors and promote positive ones.	999 high-risk men from Monrovia, Liberia were recruited. Average age 25, nearly 8 years of schooling, and a majority were involved in low skill labor and illicit work.	After one year, therapy alone led to a 0.25 SD (se 0.088) fall in antisocial behaviors, while therapy plus cash led to a 0.31 SD fall (se 0.089).	0.088 * 2.8 = 0.246 0.089 * 2.8 =0. 249	0.25 SD 0.31 SD

Barrera-Osorio et al., 2020	Vocational training programs with an emphasis on either social skills or technical skills, and a randomized stipend to cover transportation and meals costs.	663 individuals registered for the courses, 451 were assigned to training and 212 to the control group.	Vocational training increased employment by 2.16 days per month (se 1.09)	1.09 * 2.8 = 3.052	2.16 SD
Chioda et al., 2021	Skills for Effective Entrepreneurship Development (SEED) program, which includes hard skills and soft skills training.	Initially, 4,400 youth were sampled from a nationally representative sample in Uganda, with random assignment to two treatments or a control group.	Conscientiousness increased by 0.115 SD (se 0.04)	0.04 * 2.8 = 0.112	0.115 SD
Mehmood, Naseer and Chen, 2023	Deputy ministers were randomly assigned to one of the four altruism training treatment arms.	213 junior ministers, with 53 in utilitarian treatment, 54 in malleability treatment, 53 in joint utilitarian and malleability treatment and 53 in placebo.	Stand-alone Utilitarian (U) is associated with 0.5242216 SD (se 0.179515) increase in altruism.	0.179515 * 2.8 = 0.502642	0.5242216 SD

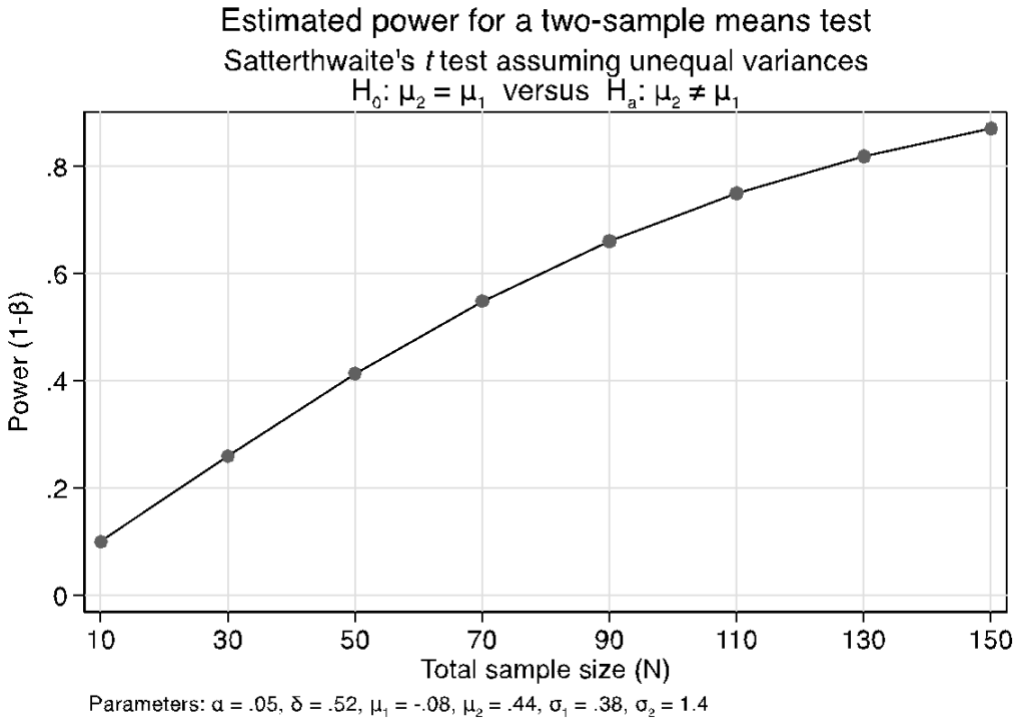
Note: Table B16 above presents the effect sizes and the ex-post MDEs from the main results across related shifting interventions. Here the MDEs are calculated according to the standard errors of the (standardized) treatment effect estimates, assuming a two-sided statistical test with a significance level of 0.05 and power of 80%. By comparing the standardized effect sizes, we can see that the effect size in our paper (0.52 SD) is the second largest among the selected studies, next to the prejudice correction effect in Devine et al. (2012).

Figure B1: Minimum Detectable Effects (MDE) across Relevant Studies



Note: The graph presents the minimum detectable effects (MDE) calculated as 2.8×SE, against the estimated treatment effects (standardized) across different studies.

Figure B2: Ex Post Calculation for the Minimum Sample Size Needed



Note: The graph presents the minimum required sample size for detecting a difference with the size of the main estimated effect (0.5242216 SD) in altruism between the treatment and control, given the desired power level. The parameters include the pre-treatment means and standard deviations of altruism for both officers in the Stand-alone Utilitarian (U) training group and those in control group.