Training Effective Altruism

Shaheen Naseer

Global Priorities Institute | July 2022

GPI Working Paper No.7 -2022

Training Effective Altruism

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

Abstract

Our randomized controlled trial of Pakistan's deputy ministers compares two schools of thought about how to cultivate prosociality. We find that training the utilitarian value of empathy results in a 0.4-0.6 standard deviation increase in altruism, cooperation, coordination, and teamwork. Field outcomes—orphanage visits, volunteering in impoverished schools, and blood donations—also roughly double. We find that treated ministers increased their mentalizing of others, both in terms of measures of theory of mind and in the field—however, blood donations only increased when their specific blood type was requested. We also find effects on language use in social media and on honesty. In contrast, we find no effects training malleability-of-the-self, even in combination with the utilitarian treatment. We interpret these results through the lens of self-image models.

JEL Codes: J01, J24; D64; D73; M50
Keywords: soft-skills, prosociality, altruism

[1]New Economic School (E-mail: smehmood@nes.ru), Lahore School of Economics (Email: ranashaheen133@gmail.com), Toulouse School of Economics (E-mail: daniel.chen@iast.fr). AEA RCT Registry number ID is: AEARCTR-0006655. A local IRB is obtained from the Lahore School of Economics with IRB Number RERC-022021-01. We would like to thank the training Academy, John List, Ray Fisman, Stefanie Stantcheva, and Martha Nussbaum for their helpful suggestions and comments. 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.

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)

I. Introduction

Prosociality—behavior that benefits others or society as a whole—is a critical concept for understanding contract enforcement, management of commons, public goods provision, establishing effective rule of law, and efficient governance (Knack and Keefer, 1997; La Porta et al., 1997; Fehr and Gächter, 2002; Ostrom et al., 2002; Henrich et al., 2004; Guiso, Sapienza, and Zingales, 2009; Bloom, and Van Reenen, 2011; Cooper and Kagel, 2015; Burks et al. 2016; Robalino and Robson, 2016; Deming, 2017). The importance of procociality to a variety of societal outcomes raises an urgent policy question: how can we cultivate prosociality? Though some laboratory studies showing short-term malleability of prosocial behavior, there have been few large-scale randomized control trials that train prosociality effectively, especially in adults. A pioneering experiment found improvements in prosociality after an early childhood intervention (Heckman et al., 2013). Building on this, recent experiments find increases in prosociality from mentoring (Falk et al., 2020) and a curriculum designed to build social cohesion (Alan et al., 2021)—both interventions were performed in schools over a single year. We demonstrate that there is an effective way to train prosociality among adults that can be scaled, which was inspired by a philosophy associated with Peter Singer (one of the "most influential ethicists alive", Goldhill, 2016). Singer proposes cultivating prosociality through utilitarianism (the principle that one should evaluate their actions by the total utility achieved, not just for them, but for society as a whole)—which he calls "effective altruism" (Singer, 2015).[2] We compare this way of cultivating prosociality to another way based on a psychological school of thought which emphasizes the malleability of the self.

[2] Peter Singer is well known for his strict adherence to utilitarianism and advocacy of animal 'liberation'. Singer's ideas are reported to inspire career choices of individuals, kidney donations, founding of large charitable, animal rights organizations and the whole "effective altruism movement" (Goldhill, 2016; Guardian, 2000, Wall Street Journal, 1999).

We study deputy ministers in an elite training Academy in Pakistan. These deputy ministers are high-stakes decision-makers who advise the President, Prime Minister, and Cabinet Ministers. This job is highly competitive, only about 1% of applicants are chosen from about 15,000. When asked, about 70% reported that their main reason for joining the public service was job perks and power, rather than a prosocial motive (Training Academy's Internal Survey, 2020). Changing this has been a key priority for the Academy.

To understand prosociality, we draw on recent economic insights into the importance of soft-skills[3] – empathy in particular (see Deming 2017). Perspective-taking—"putting oneself in another's shoes" (Premack and Woodruff, 1978)—is called "Theory of Mind" by psychologists and "Degree of Strategic Reasoning" by economists. Soft-skills have been formally modeled to reduce coordination costs enabling teams, organizations, and society can work together more effectively. There are three key challenges we face in developing an understanding of soft-skills: measuring soft-skills—such as teamwork and coordination—understanding the underlying mechanisms—such as theory of mind—and identifying causal effects (Deming and Weidmann 2021). Our paper seeks to make progress on all three.

We show that training high-stakes decision makers in the value of empathy increases their altruism, perspective-taking, and honesty. Indeed, honest public servants are important for strong governance, fiduciary duty, and rule of law. We measure perspective-taking in a competitive setting—known as the "beauty contest" or the "guessing game" (Nagel 1995)—this game is a little like rock paper scissors in that the best strategy depends on the strategy chosen by other players. High performance in these strategic dilemmas is associated with neural activity in the medial prefrontal cortex associated with successful mentalizing (Coricelli et al. 2009). We measure honesty in a die-rolling game known as the "lying game" (Abeler, et al. 2019; Gneezy, et al. 2018; Fischbacher, et al. 2013).

Besides laboratory measures of altruism—such as donations to each other and to charities—we observe field evidence consistent with effective altruism. We find that blood donations—solicited by volunteers at a prominent blood bank—increased only when deputy ministers were told that their exact blood type was in need. This is consistent with the ministers considering whether their blood donation will actually be useful. Other measures also increased— namely cooperation and coordination in strategic dilemmas, orphanage visits and volunteering at impoverished schools 4 months following the intervention, and scores on a

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

regular soft-skills assessment administered by the training Academy (which covered topics such as negotiation, cooperation and leadership). Six months after the intervention, a committee of senior public officials and former deputy ministers scored treated ministers more highly on teamwork and group decision-making assessments in a 1-day policy scenario simulation workshop.

Finally, we observe a shift in language used on social media. Treated ministers are about 20 percent more likely to use "we" rather than "I", and about 40 percent more likely to use "us" rather than "them", relative to the placebo group. This is a substantial effect—equivalent to about a doubling of the usage of "we" and "us" relative to the placebo group. Not only does our data provide novel evidence that training the utilitarian value of empathy improves theory of mind in strategic dilemmas, it is also novel in the way it links lab behavior to administrative data and field outcomes—such as blood donations, orphanage visits, volunteering, policy assessments, and social media feeds.

We compare Peter Singer's effective altruism to Carol Dweck's malleability-of-the-self—associated with a psychological school of thought on cultivating prosociality by emphasizing the malleability of empathy—and find little evidence that the latter changes our outcomes, even in combination with the former. We interpret this through the lense of the theoretical self-image models of Benabou and Tirole (2004, 2006, 2011): empathetic behavior informs one's identity as a prosocial person, but increasing the perceived malleability of prosociality makes behavior less informative about one's identity. Formally, utilitarian training increases the private benefits of empathy while malleability training reduces how much we update our self-beliefs upon taking an empathetic action. Consistent with this, deputy ministers in the malleability-of-the-self treatment thought that prosociality was less important than they did before treatment.

We demonstrate robustness of these results through a series of sensitivity analyses. First, we show the randomly assigned groups are balanced across individual characteristics, cognitive ability and pre-treatment outcomes related to prosociality. For this we use pre-treatment mathematics scores and written assessments—a measure of cognitive ability—as well as "psychological assessments"—which were utilized by a panel of psychiatrists to screen for antisocial deputy ministers who passed the written exam. We also find that our results are robust to randomization inference and adjustments for multiple outcome tests. Finally, our results are unlikely to be driven by experimental demand since 1) only those individuals whose exact blood type was requested increase their blood donations, 2) malleability treatment has no

impact on prosocial behavior, and 3) a placebo assessment of general quantitative skills shows they are unaffected.

The paper contributes to several strands of the literature. Firstly, to the best of our knowledge, we are the first to show that theory of mind (Nagel, 1995) can be altered in adults. Theory of mind is related to recognition of others—for example modelling their decisions in strategic dilemmas, understanding their point of view and emotions, and simply viewing them as having similar mental capacities to yourself. Our study is also related to the formation of prosociality (Kautz et al., 2014; Burztyn et al. 2020). A few randomized control trials find medium to long-term effects from training interventions (Heckman et al. 2013; Falk et al., 2020; Alan et al., 2021; Cappelen et al. 2020), and our results suggest the principle of effective altruism could be used as a parsimonious foundation for training interventions that increase prosociality.

Secondly, 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, 2015; Deming 2017). Soft-skills are also likely a neglected component of the personnel economics of the state. Consider that a recent literature review highlighted three important channels for improving public service in developing countries—selection, incentives, and monitoring (Finan et al., 2017)—but no attention was paid to soft-skills, nor how these "technologies" of production can be enhanced after the recruitment of public officials. To be sure, changes to selection, incentives, monitoring, or even soft-skills could theoretically decrease social welfare (Ashraf et al. 2020). However, we find evidence consistent with an increase in social welfare; for instance, emphasizing the private benefits of empathy led to increases in blood donations at a time when "blood banks were practically empty" (Shaukat Khanam Hospital, 2021).

Finally, our study applies recent theoretical developments in modeling the motivations of high-stakes decision makers such as public servants—self-image and prosocial behavior may be important drivers of effective service delivery (Besley and Ghatak, 2018; Barfoot et al., 2019; Gulzar and Khan, 2021; Ashraf et al., 2020). We map competing schools of thought on cultivating prosociality into formal models and test them empirically.

This paper is organized as follows. Section II provides background information and explains experimental set-up. Section III describes the data and empirical strategy. Section IV presents results from the lab and field outcomes—including evidence in favor of the effective altruism mechanism. Section V details a series of robustness checks. The final section concludes.

II. Background: Context and Study Design

A. *Background*

The Pakistan Federal Administrative Service inherited its structure from the Indian Public Service of British Colonial India. It is responsible for running the central administrative operations and hiring deputy ministers, who serve as key policy advisors to the President, Prime Minister and cabinet ministers. The government considers these policy advisors the "…key wheels on which the entire engine of the state runs" (Federal Government of Pakistan, 2019).

Deputy ministers are selected through competitive examinations. Initially, they must pass a written examination. Next, there is a psychological assessment—conducted by a panel of psychiatrists who analyze their "personality traits"—to determine their level of prosociality, and an interview with a panel of senior policy makers which tests their interpersonal skills and ability to work under pressure.[4] In order to be eligible to qualify for these examinations they must have completed 16 years of education or hold a bachelor's degree in any subject. Only about 200 of these 15,000 test takers are selected each year to serve as deputy ministers, making the passing percentage about 1%. The specific cohort we study had 14,521 candidates, of which 365 passed the initial written examination and 213 qualified to serve by also passing psychological and interview assessments.

Deputy ministers participate in training programs, one of which takes place at an elite training facility referred to colloquially as the Academy. Training involves participating in workshops on various subjects—for example public sector management, history, economics, politics, and professional etiquette. These public officials receive a salary of at least USD 1,000 per month depending on their seniority, as well as several perks and privileges. Specifically, the perks include free housing (a bungalow), a car, a chauffeur, a meal allowance, and domestic help. Almost 70% of them report perks and associated power as the main reason for joining the service (Training Academy Internal Survey, 2020).[5]

[4] The psychological assessment is an individual two-day-long "workshop" where each candidate, upon passing the written examination, appears before a panel of psychiatrists. They are asked to respond to images, scenarios involving vulnerable citizens and questions presented to them.

[5] As noted extensively in the literature, the associated perks are hard to value but are likely substantially larger than the USD 1,000 base salary (see e.g. Finan, Olken and Pande, 2017).

B. *Study Design*

We conduct a randomized evaluation implemented through a close partnership with the training Academy in Pakistan. The Academy is one of the most prestigious academies in the country providing training to elite policymakers. All activities at the Academy are mandatory and absences are noted on the permanent records of deputy ministers. We obtained unique access to these deputy ministers during their training and conducted a workshop entitled "Soft Skills Workshop". Our workshop was prerecorded and delivered online. Table B1 (in Appendix B) presents a flow chart of the timing, procedural details and set-up of the experiment.

Sample and Randomization.— The study took place on a cohort of 213 public officers who qualified for service the same year.[6] None of the participants had taken part in any prior randomized evaluation to the best of our knowledge. The Academy cooperated extensively before, during and after our intervention. 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).[7]

The four treatments were delivered via a non-shareable and non-downloadable link containing 4 different training lectures. The Academy explicitly prohibited the sharing of material and allowed us to designate the training as an "individual assignment". In addition, 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.[8] The training could only be accessed by entering the unique email address of the participant—which were provided to us by the Academy. Participation was mandatory for the entire cohort.

It should be emphasized that the leadership of the Academy and the Federal Government of Pakistan were very helpful. For example, the Director of the Academy sent, from his official email address, an email which said to "carefully watch the training lecture", to avoid discussing or sharing any material with their colleagues, and that "failure

[6] To protect their identity, and due to the politically sensitive nature of this experiment, we do not reveal the exact year of examination of the cohort since this could allow anyone to identify all participants of the experiment.
[7] Individual-level randomization was performed using a random number generator in Stata.
[8] We used the services of an expert computer scientist who blocked sharing and downloading of the training lecture. The COVID-19 pandemic also meant that the 213 officers were in their homes, dispersed, all over Pakistan and were not in the usual training facility in Lahore which made it even more difficult for them to discuss the material provided to them and form new social connections.

to comply may lead to disciplinary action". This email was sent to everyone, including the group receiving placebo training—only the assigned training lecture varied. To maximize comprehension and retention, we also asked deputy ministers to summarize the key lessons from the lectures in a short essay.

Table 1 reports 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.[9] Importantly, note that pre-treatment outcomes in particular those related to altruism—baseline blood donations and pre-treatment psychological assessments—are balanced. Groups are also balanced in pre-treatment measures of cognitive ability—including mathematics and written assessment scores as well as non-cognitive ability interview assessment. The similarity of all of these measures suggest the treatment groups are balanced in both individual characteristics and pre-treatment altruism.

The rollout.— The treatments were deployed on 10th October 2020 when the Academy's Director sent an email to all 213 deputy ministers. The email specified that it was part of the mandatory soft-skills workshop in their training program administered by the Federal Government of Pakistan. The email instructed them to open the assigned link associated with their name and enter their email address to access the assigned training.[10] They had a deadline of 20th October 2020 to watch the training lecture and write a short 500-word essay on the key lessons of the lecture. The link became inactive once they had finished watching the training lecture, further preventing them from being able to access materials from other treatment groups.[11] Further details on the set-up and roll-out of the experiment can be found in the flow chart presented in Table B1 of Appendix B.

Utilitarian Treatment.— Our first treatment involved the participants watching a training lecture emphasizing the value of empathy, especially how empathy can provide

[9]Following Duflo et al. (2015), Table 1 reports standard deviations in brackets and p-values corresponding to respective F statistics in italics.
[10] It is worth reiterating that the link was uniquely linked with their official email address and could not be accessed by someone else. This is possible using oTree (Chen, Schonger, and Wickens, 2016).
[11] The transcript of the email sent out to all officers is presented in Table B2 in the Online Appendix B.

private benefits to them in both personal and professional life.[12] The training reinforced this message by relying on two approaches: narratives and research studies. The lecture explains that both qualitative and quantitative evidence supports the idea that being empathic is not only prosocial but also privately beneficial. The training lecture begins by a motivating example or a "puzzle": Why do firms like Google—who aim to maximize profit—invest millions in training their employees in empathy—e.g. at the *Google Empathy Lab*—when it is costly for them? Perhaps this is a profit maximizing response on the part of Google. We go on to discuss several (truthful) real-life stories of former deputy ministers, famous for their stelar public service records, who demonstrate prosocial behavior and empathy. The training goes on to present its main findings and discuss several studies that back up these narrative accounts. For instance, we discuss studies that show that employees who demonstrate empathy benefit firms because they are better able to deal with complex social relationships and hierarchies. The training also discusses studies showing how CEOs and senior managers are more capable of motivating their employees, reducing shirking and increasing overall productivity and profits when they display more empathy specially towards their subordinates. The utilitarian training treatment concludes by reiterating its core message: "*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.*" The complete transcript of the training is presented in Table B3 of Appendix B.

Malleability Treatment.— Our second treatment group was provided with training emphasizing the malleability of empathy. That is, how empathy can change over time within a person and across populations. This treatment was inspired by prior work in psychology that indicates that the degree of empathy a person displays is not a fixed personality trait but is rather a malleable one. 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 that empathy is malleable by emphasizing qualitative and quantitative evidence. That is, this training relied on narratives of personal transformation—stories emphasizing the malleability of empathy—and quantitative research in psychology that argues both that empathy is

[12] All trainings, including the placebo, also involved the individuals writing a short 500 word essay summarizing the main lessons learned from the respective lectures.

malleable and that people can become more prosocial over time. The malleability training also concludes by reiterating its core message: "*Qualitative and quantitative evidence backs the idea that empathy is not fixed but is malleable. It is a skill that can be developed.*"[13]

Joint Utilitarian and Malleability Treatment.— Our third treatment group received a combination of the two previous treatments—they were given training that emphasized both the utility and malleability of empathy. Like our stand-alone treatments, this group received narrative accounts and quantitative evidence that indicated empathy is beneficial for them as well as malleable. This training concludes by reiterating its core message: "*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.*"[14]

Placebo.— Finally, our control group received a placebo training unrelated to the utility or malleability of empathy. They received a macroeconomics lecture taken from an undergraduate course at the Lahore School of Economics. This training lecture covered basic macroeconomic concepts—such as Gross Domestic Product, Gross National Product, Purchasing Power Parity and other macroeconomic identities. All lectures were delivered by the same person, including the placebo lecture, and every participant was tasked with writing a 500-word essay summarizing key points of the lecture.

How COVID-19 Impacted our Study Design.— Our study took place in September 2020 and deputy ministers typically reside at the Academy for their entire training. However, the cohort we studied was instructed to remain in their home cities due to the COVID-19 pandemic. The training, therefore, took place online. The Academy has strict training protocols that do not allow for random assignment by experimenters on this "elite group" of public officials. However, these procedures did not apply for off-site training, therefore, the unique circumstances provided by the COVID-19 pandemic gave us an opportunity to randomly assign training lectures to them at the individual level. A combination of factors likely reduced treatment contamination—the Academy's direct instruction that participants

[13] The complete transcript for the training lecture treatment is presented in Table B4 of Appendix B.
[14] The complete transcript for the joint utilitarian and malleability treatment is presented in Table B5 of Appendix B.

not share or discuss our workshop material with their peers, the geographical dispersion of the officers due to the pandemic, the non-shareability of the link. Although, it should be noted that treatment contamination would only indicate that our results underestimate the effect sizes.

III. Data and Empirical Strategy

A. *The Data*

The sample consists of all 213 deputy ministers who entered service in a single year.[15] The outcome variable data on behavioral games was collected during a Zoom call with everyone under supervision of the Academy in a live session. All deputy ministers participated in 12 behavioral games during the 2-hour soft-skills workshop. The administrative data on individual policy makers' characteristics, which we used in our test for balance after random assignment and as control variables in regressions, was obtained from the administrative records of the Academy. Pre-treatment blood donation data was obtained from a baseline survey. The written, interview and psychological assessment scores of the participants were obtained from the Federal Public Service Commission (FPSC) of Pakistan, who oversee and organizes these assessments.[16] Finally, field outcomes related to blood donations were obtained from a prominent blood bank; we worked closely with volunteers requesting blood donations at the bank.[17]

Outcome Variables.— Our 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).[18] The decision of the dictator is our first measure of altruism—we use this because the game holds for many real world settings of

[15] The year is anonymized on request of the Academy citing legal and political concerns.
[16] The Federal Public Service Commission (FPSC) is a statutory body of the Government of Pakistan, constituted in 1947. It obtains its jurisdiction from the Constitution of Pakistan and its responsibilities include recruiting elite policy advisors and administering their entry examinations and assessments.
[17] An IRB was obtained where the experiment was approved by Lahore School of Evonomics's Ethical Review Board (IRB) who approved the IRB after close coordination and consultation with the Academy officials.
[18] 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.

altruistic behavior (Henrich et al., 2005; Levitt and List, 2007; Kosse et al., 2020).[19] Our dictator game results are interesting since, rather than studying students who have self-selected for the experiment, we administer these games with deputy ministers who have not self-selected to be a part of a lab experiment—hence we present important new work that moves beyond student populations (see e.g. Cappelen et al., 2015).

Our second measure is a 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 told 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 behavioral games are normalized between 0 and 1 to make comparisons across games easier. In Appendix B, we also present results for outcome variables standardized to mean zero and standard deviation one.

Finally, we assess prosociality in the field. We randomized participants in each group, then collaborated with the volunteer group of a prominent blood bank in Lahore to design the script they would use to call each deputy minister with a truthful and urgent request to donate blood.[20] We measure outcomes for the public servants agreeing verbally to donate blood, as well as those who booked a definite appointment to donate blood during the call. The phone calls requesting blood donations took place about 2 months following the roll-out of our training lectures and submission of the summary. Using a unique dataset from a COVID-19 survey with the Academy, we also utilize information on the actual blood group of these deputy ministers and randomly assign participants in each treatment arm to a group where we urgently request their exact blood type, while the remaining individuals within each treatment arm are randomly assigned an urgent generic request for blood donation but without explicit mention of the blood type of the deputy ministers.[21]

Likewise, two syndicate field trips took place about four months following the training. For the first field trip the deputy ministers had to choose between attending a lecture by a senior

[19] 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..

[20] The shortage of blood meant that this was truthful information since all blood types were urgently needed since there was a steep fall in blood donations following the COVID-19 pandemic. According to one of the volunteers making the calls: "the blood banks were practically empty".

[21] 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.

bureaucrat or visiting an orphanage. For the second trip, about 6 months following the treatment, deputy ministers had to choose between volunteering at an impoverished school selected from a particular government network of schools or attending a lecture by a senior bureaucrat. The Academy shared this data with us, and we use it as a field-based measure of altruism or prosociality. Other field-based measures include their choice of a book on empathy in a book lottery elicited at the end of the 2-hour soft-skills workshop, their use of terms associated with social cohesion on social media, and their grades on both soft-skills and teamwork assessments. This soft-skills workshop tested material related to negotiations, leadership, teamwork and cooperation. The teamwork workshop was scored by a panel of senior bureaucrats, policymakers and academics and involved deciding on policy responses as part of a team. For instance, this scenario question was 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?*" (FPSC, 2021).

Explanatory Variables.— Our main treatment variables are dummy variables for the three treatments. U_i and M_i are binary dummy variables for the stand-alone utilitarian and the stand-alone malleability treatment arms, respectably. UM_i indicates the joint utilitarian and malleability treatment arm. We control for all individual characteristics available from administrative data—specifically this includes written, mathematics, psychological and interview assessment scores from the 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.

B. *Attrition*

Due to close cooperation with the Academy and the fact that our workshop was compulsory, 100% of the cohort undertook the treatment. There was, nevertheless, some attrition in recording the blood donation outcomes. That is, when the blood bank called the deputy ministers, requesting for the blood donation, some did not pick up the phone or refused to give an answer on the blood donation request.[22] However, given the prominence and credibility of the blood bank, only 8 out of the 213 did not respond to the call made by the

[22] Most "non-respondents" requested the blood bank to call them back but never picked up the phone again. We report the most conservative estimates excluding these public officials although coding these individuals as "no" increases the sample size and precision of our estimates.

blood bank. Roughly 95% of participants gave a definite response both to whether they would donate blood and to whether they would set a definite appointment with the blood bank. We do, however, show that there is no evidence for differential attrition for either agreeing to donate blood or setting up a definite appointment for the blood donation (these results are reported in Table B6 of Appendix B).

C. Estimation Strategy

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

$$Y_i = \alpha + \beta U_i + \gamma M_i + \delta UM_i + \boldsymbol{X}_i' \mu + \epsilon_i \quad (1)$$

where Y_i is the 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; $\boldsymbol{X}_i'$ 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 of the 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 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 $\gamma = \beta + \delta$). We report ordinary least squares (OLS) estimations. The results are qualitatively unchanged with probit or logit estimations for binary outcomes. The results are also unchanged when we add a large number of individual level controls (these include scores on pretreatment written, interview, mathematics and psychological assessment scores, asset ownership, income, age, years of education and dummies for foreign visits, gender, birth in political capitals and professional designation).

IV. Results

A. *Results from Dictator Games*

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 increases altruism. Since we have normalized the outcome variable to be between 0 and 1, the table shows that the utilitarian treatment increases altruism by about 6 percentage points—equivalent to an approximately 12% increase over the placebo mean. The coefficient estimates are similar without controls for a large number of individual level characteristics. Likewise, in Table 2, we also report results for an oft-used variant of the dictator game where donations to UNICEF charity are solicited instead of donations to strangers. The effects, reported in columns (3) and (4) of Table 2, are even larger. The utilitarian treatment increases in altruism by about 20 percentage points—equivalent to an approximately 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 reported in Table B7 of Appendix B where we standardized the outcome variables to mean zero and standard deviation one. For a reference point, the effect sizes of our utilitarian training intervention (video lecture and summary of the lecture) 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).

Next, we investigate if whether our treatment effects whether deputy ministers want to learn more about empathy. We record revealed preference by offering to send one of two books to each deputy minister at the end of the workshop. The first book is on empathy (*Mindsight: Transform Your Brain with the New Science of Empathy* by Daniel J. Siegel) and the other is our placebo book, a book on basic econometrics (*Mastering Metrics* by Joshua Angrist and Jörn-Steffen Pischke). Figure 1 as well as the first two columns of Table 2 present these results across our randomly assigned groups. We find the group allocated the stand-alone utilitarian treatment is about 20 percentage points more likely to choose the book on empathy relative to the placebo group—equivalent to an approximately 50% increase over the placebo mean. This suggests that our training induced the utilitarian group to not only respond differently in dictator games but also to become more curious about empathy.

These results are particularly interesting when combined with the scores on a regular soft-skills assessment organized by the Academy at the end of each training program. The

workshop assesses these policymakers on soft-skills. This includes policy scenario assessments related to negotiations, leadership, teamwork, and cooperation in public policy making. In Table 3 (column 3 and 4), we observe that the group that received the utilitarian treatment scores about 10 percentage points higher than the placebo group, a 20% increase over the placebo mean. Taken together, the results from Table 3 indicate that our treatment not only increased curiosity to learn more about soft-skills but also likely affected the investment of effort to learn these skills—as revealed through higher scores in the regular soft-skills assessment.

The question remains—do our results from the dictator games and revealed preference measures change real-world altruistic behavior? Both Henrich et al. (2005) and List and Levitt (2007) have noted that several reasons, ranging from culture and environment to self-selection of experimental subjects, make extrapolating altruism in behavioral games to real-world behavior difficult. In the next subsection, we provide evidence of empathetic behavior from the field.

B. Results from the Field

We use data on the blood groups of the deputy ministers and randomized phone calls to measure altruism in 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 type was in urgent need, while the other half (107 participants) received an urgent request to donate blood but without any mention of blood type. That is, the first group was told that "O Positive Blood is urgently needed"—where the deputy minister had the O Positive blood group—and the second group was told only that "Blood is urgently needed". These requests for blood donations are made about 1.5 months after the training.

The first two columns of Table 4 reports results related to whether participants agreed to donate blood, while the latter two columns report results related to setting up a definite appointment for blood donation at the bank. The estimates in column (1) and column (3) reveal the large effects of the utilitarian treatment: the stand-alone utilitarian group is about 25 percentage points more likely to both agree to donate blood and to set up a definite appointment with a blood bank relative to the group that received the placebo training. This is a substantial effect—equivalent to an increase of about 80% compared to 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 (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 significant effect on blood donations relative to placebo group, which is consistent with our results from the dictator games, empathy book choice, and soft-skills assessment.

This, however, masks important heterogeneity among those that were randomized into the group that were requested that their exact blood group was in need relative to those that were made a generic request for blood donation. Columns (2) and (4) of Table 4 report estimates of the interaction of terms for our 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 on blood donations seems to be *entirely* explained by the utilitarian group's reaction to the request for their exact blood type was needed.

This result can be observed most clearly in Figure 3: we observe that the blood donations for the utilitarian group more than doubled when their specific blood type was requested (left panel). We do not, however, 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. These results indicate that utilitarian deputy ministers are "effective altruists" (Singer, 2015; MacAskill, 2019)—they respond altruistically only when they believe that their blood is likely to be helpful.

Finally, we obtained data from the Academy on regular "syndicate field trips" that they undertook 4 and 6 months following the treatment. The deputy ministers were given the option by the Academy to either visit a prominent orphanage (*Dar-ul-Aman*) or attend lectures about a specific government program delivered by a "veteran" policy official. This data was collected separately from the research team and so is unlikely to be affected by experimenter demand. 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 rather than attend the lecture (Table 4, Column 5). This is equivalent an approximately 80% increase over the placebo mean. These results are corroborated by the results from a second field trip 6 months after the treatment: the deputy ministers were given the choice between volunteering to teach for a week in any impoverished government school that falls under the Progressive Education Network (PEN) or to, once again, 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 in at the

PEN education network's impoverished schools, significantly increasing the rate of volunteering over the placebo mean.

Substantively, these results 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 4 and 6 months after our trainings, and (2) these data come directly from the Academy and are part of their regular training curriculum, providing an external corroboration of our results.

C. *Impact on Perceived Importance of Emotional Intelligence and Teamwork*

Since deputy ministers undertake different job designations and act as key advisors to top public officials, it is challenging to assess and compare their performance as policy advisors directly. We use two variables as proxies for their performance and decision-making process. First, we assess how important they perceive Emotional Intelligence to be for policy-making. Second, we leverage a unique teamwork policy simulation assessment that is used to gauge their performance at the Academy. We also have available a placebo outcome—assessment of quantitative ability—that took place at almost the same time as the teamwork assessment. Therefore, next, we investigate the impact of our treatment on perceived importance of Emotional Intelligence, regular teamwork and quantitative assessments conducted by the Academy. We find this interesting because this data was collected 6 months after the intervention. Columns (1) and (2) of Table 5 report these results, both with and without controls. We find that individuals in the stand-alone utilitarian treatment give a 0.8 standard deviation higher rating to importance of emotional intelligence in policy making. This is equivalent to a one-point increase on a 5-point scale and suggests that our training had a large and potentially long-run impact on how important policy makers regard soft-skills.

Our results on skills pertaining to effective teamwork are also likely to hold in the field. Deming and Weidmann (2021) have given us important new research which shows that teamwork—a soft skill—explains several puzzles in the labor market. We utilized a day-long teamwork workshop involving simulated policy scenarios—assessed by a panel of experts. In this teamwork workshop, ministers were assigned to groups with four subordinates and given concrete scenarios. A typical scenario question is as follows: "*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?*" (FSPC, 2021).

The responses were scored by a panel of experts including former top officials (former supreme court judges, prominent academics, former senior deputy ministers). It was also high-stakes for the deputy ministers, since the result determined promotions and transfers. Columns (3) and (4) of Table 5 present these results: we observe that stand-alone utilitarian groups have about 0.6 standard deviation higher scores on these teamwork policy assessments relative to the placebo group. We also find no evidence of the malleability treatment impacting teamwork. Reassuringly, we also find no effect of our treatments on scores of the regular quantitative assessment that took place around the same time—this serves as an important placebo check since nothing in our treatment emphasized quantitative skills.[23] These results strongly suggest that the utilitarian treatment has a real impact on soft-skills which is persistent for at least 6 months following the treatment.

D. *Impact on Social Media Feeds*

Several influential psychology studies argue that ingroup language of "us versus them" and "we versus I" are key determinants of social cohesion and that use of such language is highly transmissible and perpetuate discrimination (see e.g. the classic study by Maass et al. 1989). Recent work also finds that social media messages of a country's thought leaders can affect the beliefs and knowledge of their followers (Atlas et al., 2019). In this subsection, we examine the impacts of our treatments on use of language in social media. Out of the 213 deputy ministers, we were able to match 98 of them to their social media feeds—using their full name and file photos—prior to the onset of the training program.[24] The ministers are evenly spread across our treatment groups with 20 in Utilitarian, 30 in malleability, 20 in joint and 28 in the placebo group. Broadly, we find language indicating social cohesion increases in the utilitarian group. We also observe no impact of the malleability and joint treatments. Figure 4 reports these results: we observe that the utilitarian group is twice as likely to use "we" relative to "I" and more than twice as likely to use "us" relative to "them" in their post-treatment social media posts 5 months following the intervention.

In Table B10 (Appendix B) we present these results in regression-table-form with controls, with a standardized mean of zero and standard deviation of one. Considering the

[23] The quantitative assessment is also called a research methods assessment and was recently introduced in 2017. This tests policymakers in basic hypothesis testing and simple regression framework with applications to policy questions.

[24] These are deputy minister's feeds from Facebook, which is the most prominent social media platform in Pakistan.

sample size, the results are unsurprisingly imprecise but largely consistent with other evidence: the utilitarian treatment increases the use of "we versus I" by about 0.5 standard deviations and use of "us versus them" by about a full standard deviation. These results are suggestive of a change in the use of language associated with social cohesion for our stand-alone utilitarian group. That is, the individuals assigned the utilitarian treatment increase their altruistic behavior—as measured by blood donations, volunteering, and orphanage visits—and also use language that displays higher regard for others on social media 5 months after our intervention.

E. *Exploratory Analysis of Mechanisms*

The results so far show that the utilitarian treatment boosts altruism, teamwork, perceived importance of emotional intelligence and outcomes related to successful mentalizing of others. This subsection contains evidence that the utilitarian treatment also increases other soft-skills measured using the tools of experimental economics—in particular, estimates of the impact of our treatment on cooperation, coordination, and theory of mind are presented in Table 6 (see Sutter et al., 2019 for a review of these measures).

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. Deviating is a safe option as it 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 to real-world team behavior (Grossman and Baldassarri 2012; Barr and Serneels 2009).

In Columns (1) and (2) of Table 6 we observe individuals receiving the stand-alone utilitarian treatment perform better in the cooperation game. Specifically, they score 14 percentage points higher than the placebo group. Likewise, in Columns (3) and (4), we find they also perform better in the coordination game—the group receiving the utilitarian treatment score about 7 percentage points higher than the Nash equilibrium of this game. This is

equivalent to a 0.4 standard deviation higher score for the utilitarian treatment on decision-making and coordination.[25] Importantly, this suggests that cooperation and coordination, rather than simply redistributive preferences, drive the behavioral changes. This is an important distinction given that highly skilled "cognitive" occupations are increasingly valuing soft-skills surrounding teamwork to enhance productivity (Deming, 2017).

Human interaction also requires a capacity that psychologists call theory of mind—the ability to attribute mental states to others based on their behavior, or more colloquially to "put oneself into another's shoes". We estimate the decision-maker's theory of mind using a guessing game. In this game, each decision-maker in a group submits a number between 0 and 100. The average of the numbers, divided in half, is the target number. The decision-maker whose guess is closest to the target number wins (Nagel 1995). This is intuitively similar to a rock-paper-scissors game where players must mentalize and predict other's actions.

The results of the guessing decision-making game are reported in Columns (5) and (6) of Table 6. We find that utilitarian treatment raises the probability of being the most accurate guesser by about 10 percentage points. That is, 20% of those in the utilitarian treatment won the guessing game, which is significantly higher than about 12% in the malleability and joint treatment, and much higher than the 9% in the placebo treatment.[26] This suggests our treatment was successful in increasing altruism via increasing theory of mind in participants. These results are also 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—the final game measures lying. Each player rolls a 6-sided die and is asked to report the outcome, with higher rolls receiving higher payoffs. There is an incentive to lie rather than truthfully reveal the outcome of the die roll. That is, deputy ministers have the option of winning dishonestly by misreporting (see Fischbacher, et al., 2013; Gneezy, et al., 2018; Barfort et al. 2019). Figure 5 presents the results of the lying game. We find that the utilitarian group is significantly less likely to lie in this 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. These results are also consistent with a mechanism of effective altruism. In particular, they might be acting like homo kantiensis—a term coined by Alger and

[25] The standardized equivalent to Table 6 where dependent variables are standardized to mean zero and standard deviation 1 can be found in Table B8 of Appendix B.
[26] 9% is the mean dependent variable in Columns (5) and (6). Adding the coefficient on the utilitarian treatment yields 20%. Adding the other treatment coefficients with the placebo yields 12%.

Weibull (2013) for actors who decide by asking "would this decision best for society if everyone were to do it?"

While we hypothesize that successful mentalizing of others is likely to be the mechanism behind our key results, we investigate and rule out alternative mechanisms such as redistributive preferences and competitiveness. Consider this alternative mechanism: the utilitarian treated group might have become more competitive, donating blood 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 whether their specific blood type was mentioned. Alternative mechanisms also include public officials becoming more patient, or more trusting, or to prefer redistribution. Six other games provide evidence against alternative mechanisms that may explain our findings—Table 7 reports these results.[27] We find no effect of any of our treatments on competitiveness, patience, perseverance, redistribution, risk, or trust games (Berg et al., 1995; Fisman et al., 2007; Barling and Fishbacher, 2012; Dohmen et al., 2018; Bašić et al., 2020; Falk et al., 2020).

This exploratory analysis of mechanisms is also summarized in Figure 6 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, guessing, 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 form a common mechanism responsible for the treatment effects we estimate, while changes to patience, perseverance, redistribution, risk preferences, or trust are unlikely to be driving the results. The results, therefore, paint a consistent picture that treated deputy ministers are likely "effective altruists" in that they donate blood when it is most likely to be utilized, and possess improved theory of mind, ability to coordinate and ability to cooperate.

F. *Discussion*

The framework of self-image models from Benabou and Tirole (2011) put the utilitarian and malleability treatment in contrast. A typical model gives agents a payoff function such that: $U(a) = (v + y)a + \mu E(x \mid a)$, where v is prosocial identity, y is extrinsic payoffs, and $E(x \mid a)$ the perception of prosocial identity. The first term captures intrinsic motivation, for example,

[27] The null results are essentially identical if we standardize the dependent variable instead of normalizing it. See Table B9 in Appendix B for these results.

hardwired altruism. The utilitarian treatment trains altruism, so it may increase the intrinsic motivation to act in a way that benefits others. The second term captures extrinsic motivations to perform prosocial behaviors. The malleability training is likely lower the payoff from updating perceptions of prosocial identity because it argues that prosocial behaviour is malleable—overall, this could lower the utility gained from acting prosocially. Indeed, in Appendix Table B11, we find that deputy ministers trained with the malleability-of-the-self treatment rate prosocial traits as being less important. Deputy ministers trained with both utilitarian and malleability treatments are unaffected, perhaps in part because the utilitarian treatment emphasized private benefits of empathy (represented by y in the model).

This framework suggests that when both treatments are implemented jointly, people have greater incentive to act empathetically because of the benefits of empathy, but *also* have reduced incentive to be empathetic in any given decision since the decision does not affect their perceptions of prosocial identity. We interpret this as a reduction of μ parameter within the self-image framework—that is, the deputy ministers put less weight on updating perceptions upon taking actions, after the malleability of self or the joint treatment. As a result, the joint treatment may have qualitatively different effects than would be suggested by considered the effects of each treatment separately.

V. Robustness

Balance.— Earlier, we observed that the sample is balanced across a host of individual characteristics: income, age, years of education, gender, birth in political capitals, asset ownership, and foreign visits. It is important to emphasize that the large effects we observe are also unlikely to result from lack of balance in altruism or some ability of the deputy ministers. The rich set of outcome variables data gives us access to several pre-treatment outcomes including proxies for baseline altruism. For instance, baseline blood donations are balanced across the treatments, as are scores on psychological, written, and mathematical tests, as well as interview assessments—together this indicates that the candidates are balanced in terms of underlying ability. Important to note is that deputy ministers are balanced in terms of psychological assessments—which explicitly attempt to screen out those with low levels of prosociality—and baseline blood donations.

Sample Size and Statistical Power.— The focus on deputy ministers that make high-impact policy decisions allows us to study an elite group of high-stake decision-makers who can potentially impact long-run economic development. However, studying such a specific group necessarily made our sample sizes relatively small. Our sample of around 200 deputy ministers might raise concerns about lack of statistical power. However, our power calculation with statistical power 80% and significance level 5% reveals that even with 50 individuals per cluster, the individual level randomization allows us to detect a minimum detectable effect equivalent to a change of 0.27 standard deviations. Fortunately, our documented effect sizes are about twice as large as this, so we were able to detect the effects with our sample (see for instance the standardized results of dictator and charity game in Table B7 of Appendix B). Nevertheless, Imbens and Rubin (2015) recommend that small sample randomized trials conduct randomization inference—a process in which the econometrician scrambles the data, re-assigning treatments, and then compares the distribution of placebo estimates with the estimate from the experiment. We report in Table B12 of Appendix B the corresponding p-values with 1000 iterations of this process.[28] 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.

External Validity.— As List (2020) notes, "all results are externally valid to some setting, and no result will be externally valid to all settings." Therefore, we follow the List (2020)'s SANS (*Selection-Attrition-Naturalness-Scaling*) conditions in our discussion of generalizability of our results. First, in terms of selection, our sample consists of all 213 elite policymakers that entered service in Pakistan via competitive examinations in an anonymized year. In behavioral games our compliance is 100% given our close cooperation with the Academy, while in blood donations, volunteering, orphanage visits we still have close to 90% compliance given the credibility of prominent blood bank soliciting calls and the Academy organizing the field visits. Considering the naturalness of the setting, time frame and choice task, we obtain natural measures such as blood donations. The deputy ministers perform several natural tasks in the field especially blood donations, teamwork assessment, volunteering and orphanage visits. Finally, in terms of scaling our intervention to increase effective altruism in other settings, the intervention is cheap to deliver and may be particularly useful for developing

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

countries who face strict resource constraints. The soft-skills training is delivered online so may also be scaled to other high-stake decision makers such as judges and CEOs in several developing countries. We, however, view these results as a WAVE1 insight, in the nomenclature of List (2020), replications need to be completed to understand if the effect sizes can be applied to other general populations or high-stake decision makers in other contexts.

Multiple Hypothesis Testing. — Another key empirical issue is that we are testing multiple hypotheses. The elite public officials played 12 games and were assigned to three treatment arms, so we conducted 36 hypothesis tests. Under the assumption that none of the treatments have any effect on any outcome (all null hypotheses are true), and that the outcomes are independent, then the probability of one or more false rejections when using a critical value of 0.05 is 1-0.95^36 = 85%. As a result, in order to reduce the likelihood of these false rejections, we adjust for the fact that we are testing for multiple hypotheses. Following the literature, we use sharpened False Discovery Rate (FDR) q-values suggested by Anderson et al., 2008 (see for instance Heckman et al., 2018 for an application). These sharpened q-values are presented in square brackets in Table B13 where we also show standard p-values from our regressions in parentheses for comparison. Similar results are found when we employ List el al., (2019) familywise error rate correction (FWER) that uses a bootstrapping approach to incorporate the point dependence structure of different treatments and also allows p-values to be correlated while adjusting for multiple hypotheses (List et al., 2019). Our results remain robust at conventional significance levels.

Experimental Demand.— It is also unlikely that experimental demand drives our results—experimental demand would be driving our results if our effects were driven by participants behaving the way they feel they are expected to by the experimenter. This is unlikely for 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, enhanced theory of mind for the utilitarian group is challenging to explain through experimenter demand since it is arguably a difficult task. Thirdly, a number of high-stakes[29] administrative assessment scores including soft-skills and teamwork assessments were

[29] These assessments are extremely high stakes since they determine promotion and transfers of the deputy ministers.

conducted separately from the research team as part of regular coursework for the Academy.[30] Taken together, the measurements and patterns of data in our experiment make experimenter demand unlikely. Finally, we only found strong evidence (robust to multiple hypothesis testing, experimenter demand, small samples, and lack of balance) for the utilitarian treatment, if experimental demand was driving our results we would expect to find an effect for some of the other treatments.

VI. Conclusion

We tested different schools of thought about cultivating prosociality, finding significant impacts only from training the utility of empathy. Soft-skills have been formally modeled to reduce coordination costs, allowing teams, organizations, and even whole societies to work together more effectively. We provide causal evidence that training effective altruism impacts soft-skills such as teamwork and coordination, as well as theory of mind (mentalizing)—which is critical in models of soft-skills.

Laboratory measures of altruism, charitable donations, cooperation, coordination, honesty, and theory of mind in strategic dilemmas were all impacted. These effects are persistent over the next six months. Treated ministers donated blood twice as often in response to a request from a prominent blood bank—but only when their specific blood type was in need. Orphanage visits and volunteering in impoverished schools also increased. Test scores on soft-skills and independent assessments of teamwork and group-decision both increased by 0.8 of a standard deviation. Training effective altruism has a similar effect size on prosocial behavior (0.4-0.6 of a standard deviation) as a one-year mentoring program of elementary school children (see Falk et al., 2020).

Much of the literature has focused on childhood interventions. Although some work on workplace-based programs teaching character skills exist, no randomized control trial has previously tested different schools of thought about training prosociality in adults (Kautz et al., 2014). We show that empathy can be enhanced among even adults, which is consistent with evidence that the adult brain continues to be plastic (Duffau, 2014) and evidence that cognitive behavioral therapy can impact adults in Liberia (Blattman et al., 2017).

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

The estimated effects on perspective-taking from a recent one year 3 hours per week curriculum found an effect size of 0.3 standard deviations for Reading the Eyes in the Mind Test for school children (Alan et al., 2021). In this test, subjects are asked to guess the emotion from a pair of eyes. The guessing game (Nagel, 1995) is a strategic dilemma and also measures the ability to take the perspective of others. The estimated effect in the guessing game in our intervention is about twice as large, i.e. 0.6 standard deviations. Language used on social media is also impacted.

The utilitarian intervention is both time-efficient and cost-effective, especially because it can be delivered online. Implementing this intervention presents negligible opportunity cost of time compared to any randomized control trial tested intervention that impacts prosociality. The principles of effective altruism may be an organizing theory for effective cohesion policies amid fragility, corruption, conflict, and violence. Indeed, future research could test additional schools of thought on normative ethics besides the two in our study.

References

Abeler, J., Nosenzo, D. and Raymond, C. 2019, Preferences for Truth-Telling. Econometrica, 87: 1115-1153.

Alan, S., Baysan, C., Gumren, M. and Kubilay, E., 2020. Building social cohesion in ethnically mixed schools: An intervention on perspective taking. Accepted at Quarterly Journal of Economics.

Alatas, Vivi, Arun G. 2019. Chandrasekhar, Markus Mobius, Benjamin A. Olken, and Cindy Paladines. When celebrities speak: A nationwide Twitter experiment promoting vaccination in Indonesia. No. w25589. National Bureau of Economic Research, 2019.

Alger, I. and Weibull, J.W. 2013., Homo Moralis—Preference Evolution Under Incomplete Information and Assortative Matching. Econometrica, 81: 2269-2302.

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.

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.

Athey, S. and Imbens, G.W., 2017. The econometrics of randomized experiments. In Handbook of economic field experiments (Vol. 1, pp. 73-140). North-Holland.

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.

Bartling, B. and Fischbacher, U., 2012. Shifting the blame: On delegation and responsibility. The Review of Economic Studies, 79(1), pp.67-87.

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.

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.

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.

Barr, Abigail, and Pieter Serneels. "Reciprocity in the workplace." *Experimental Economics* 12.1 (2009): 99-112.

Besley, Timothy, and Maitreesh Ghatak. "Prosocial motivation and incentives." *Annual Review of Economics* 10 (2018): 411-438.

Bloom, N. and Van Reenen, J., 2011. Human resource management and productivity. In Handbook of labor economics (Vol. 4, pp. 1697-1767). Elsevier.

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.

Cappelen, A, List, J. Samek, A. and Tungodden, B.. "The effect of early-childhood education on social preferences." Journal of Political Economy 128, no. 7 (2020): 2739-2758.

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.

Eckel, C.C. and Grossman, P.J., 2000. Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. Experimental Economics, 3(2), pp.107-120.

Deming, D.J., 2017. The growing importance of social skills in the labor market. The Quarterly Journal of Economics, 132(4), pp.1593-1640.

Deming, D. and Weidmann, B., 2021 . Team Players: How Social Skills Improve Team Performance. Forthcoming Econometrica.

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.

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.

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.

Engel, C., 2011. Dictator games: A meta study. Experimental economics, 14(4), pp.583-610.

Güth, W., Schmittberger, R. and Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of economic behavior & organization, 3(4), pp.367-388.

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.

Finan, F., Olken, B.A. and Pande, R., 2017. The personnel economics of the developing state. In Handbook of Economic Field Experiments (Vol. 2, pp. 467-514). North-Holland.

Fischbacher, Urs, and Franziska Föllmi-Heusi. "Lies in disguise—an experimental study on cheating." *Journal of the European Economic Association* 11.3 (2013): 525-547.

FPSC (2021). Federal Public Service Commission of Pakistan. Administrative Data Archives.

Goldhill, O (2016), "The most influential ethicist alive says the world is actually becoming a better place", Quartz. Retrieved on July 3, 2021.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel. 2018. "Lying Aversion and the Size of the Lie." *American Economic Review*, 108 (2): 419-53.

Grossman, Guy, and Delia Baldassarri. "The impact of elections on cooperation: Evidence from a lab-in-the-field experiment in Uganda." *American journal of political science* 56.4 (2012): 964-985.

Guardian (2000) The most dangerous man in the world. Retrieved from: https://www.theguardian.com/lifeandstyle/1999/nov/06/weekend.kevintoolis

Gulzar, S., and Khan, Y. (2021) Good Politicians: Experimental Evidence on Motivation fo Political Candency and Government Performance. Mimeo.

Hanna, Rema, and Shing-Yi Wang. 2017. "Dishonesty and Selection into Public Service: Evidence from India." *American Economic Journal: Economic Policy*, 9 (3): 262-90.

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.

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.

Henrich, N., Ensminger, J., Tracer, D., Marlow, F., Patton, J., Alvard, M., Gil-White, F. and., Barr, A. 2005. Economic Man in cross-cultural perspective: Ethnography and experiments from 15 small-scale societies. Behavioral and Brain Sciences, 28(6), p.795815.

Imbens, G.W. and Rubin, D.B., 2015. Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.

Kahneman, D., Knetsch, J.L. and Thaler, R.H., 1986. Fairness and the assumptions of economics. Journal of business, pp.S285-S300.

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.

Fisman, R., Kariv, S. and Markovits, D., 2007. Individual preferences for giving. American Economic Review, 97(5), pp.1858-1876.

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.

List, J.A., Shaikh, A.M. and Xu, Y., 2019. Multiple hypothesis testing in experimental economics. Experimental Economics, 22(4), pp.773-793.

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.

MacAskill, W., 2019. The Definition of Effective Altruism. Effective Altruism: Philosophical Issues, 2016(7), p.10.

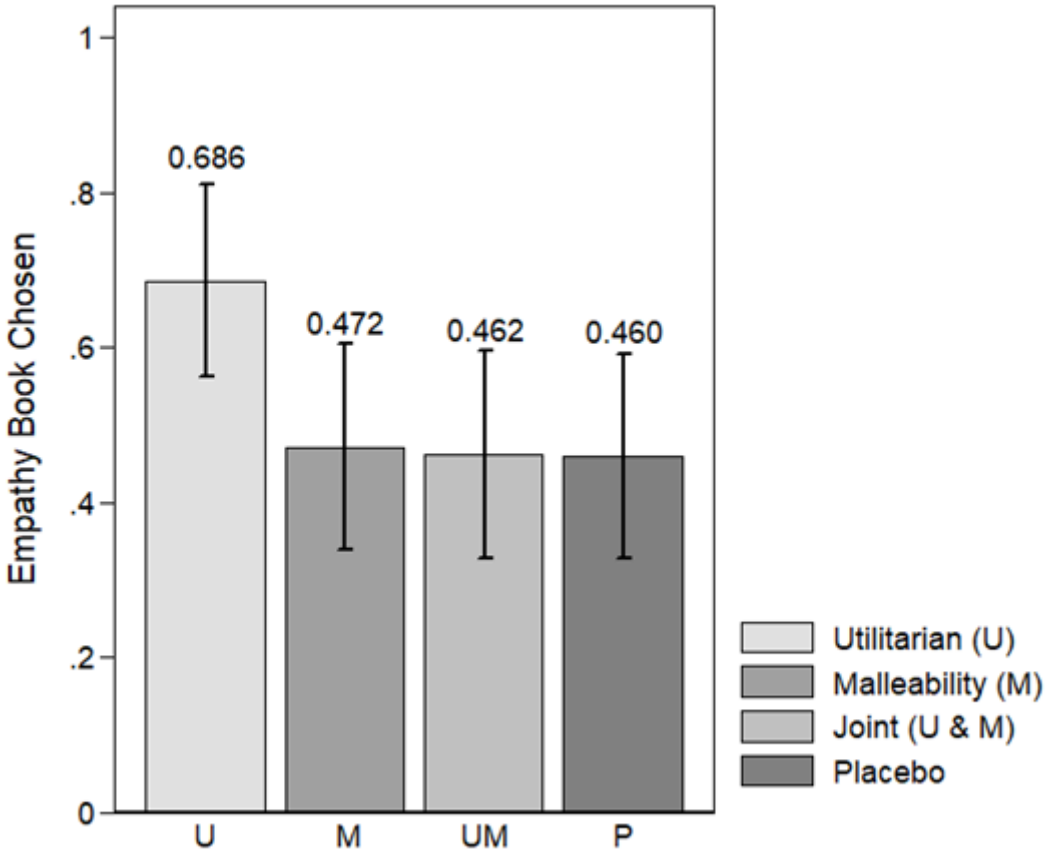
Maass, A., Salvi, D., Arcuri, L. and Semin, G.R., 1989. Language use in intergroup contexts: The linguistic intergroup bias. Journal of personality and social psychology, 57(6), p.981.

Nagel, R., 1995. Unraveling in guessing games: An experimental study. The American Economic Review, 85(5), pp.1313-1326.

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

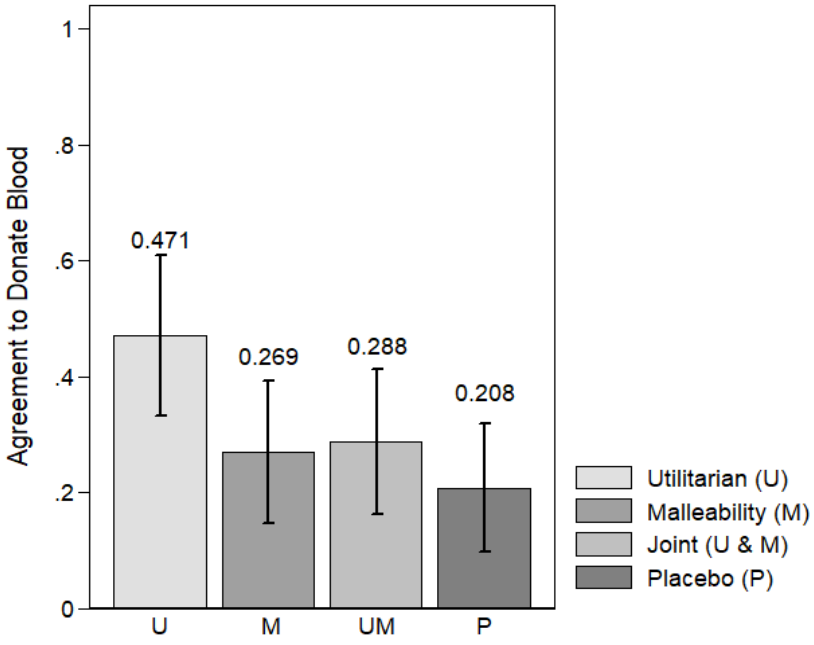
Figure 1: Book on Empathy



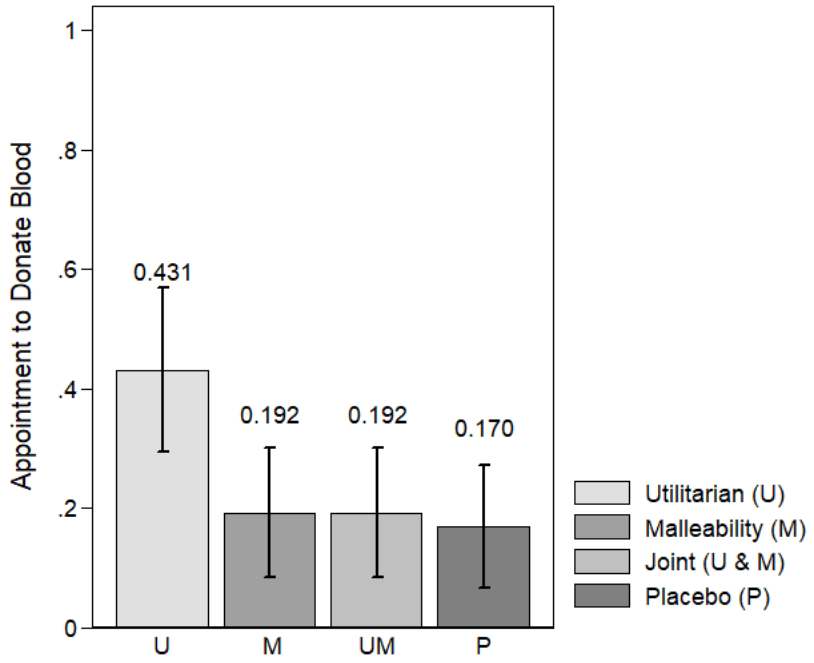
Note: The figure provides averages for the four randomly assigned groups along with the associated confidence intervals. Each bar reports the average fraction of people who selected the book on empathy according to the randomly assigned group.

Figure 2: Impact on Blood Donations

Panel A: Agreement to Donate Blood

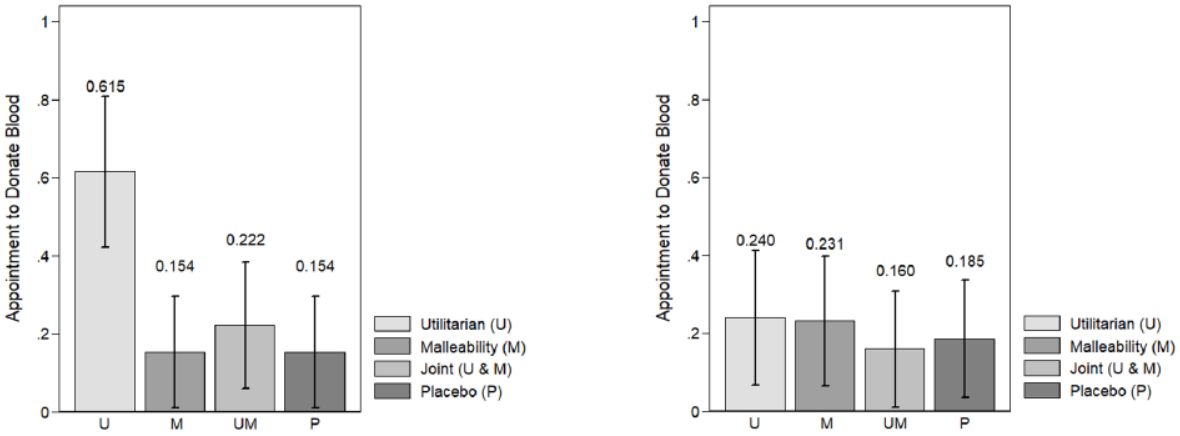


Panel B: Appointment to Donate Blood



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

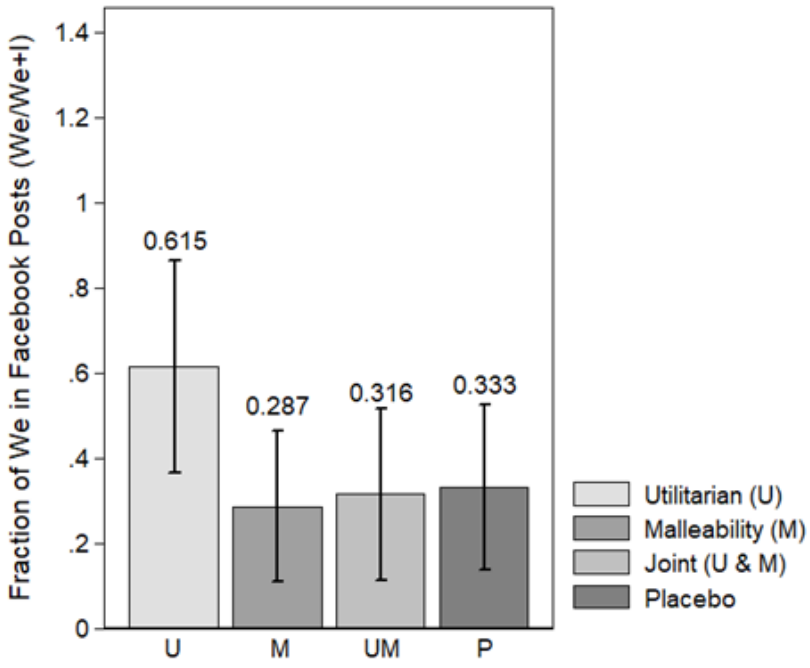
Figure 3: Impact on Blood Donations by specific versus generic request



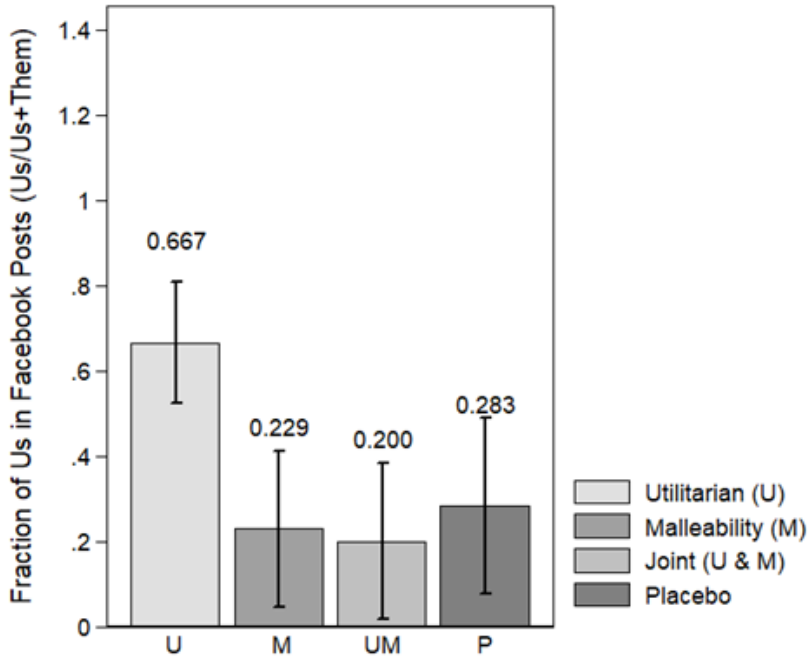
Note: The figure above provides averages for the four randomly assigned groups along with the associated confidence intervals. The figure on the left presents results on urgent truthful requests to donate blood with specific matching blood type of the individual, i.e., "O Positive Blood is urgently needed" (where the individual had the O Positive blood group). The figure on the right report results from a generic request to donate blood i.e. "Blood is urgently needed". These requests for blood donations were made 1.5 months after the intervention by volunteers at a prominent blood bank.

Figure 4: Language Use in Social Media

Panel A: Effect on fraction of "we vs I"

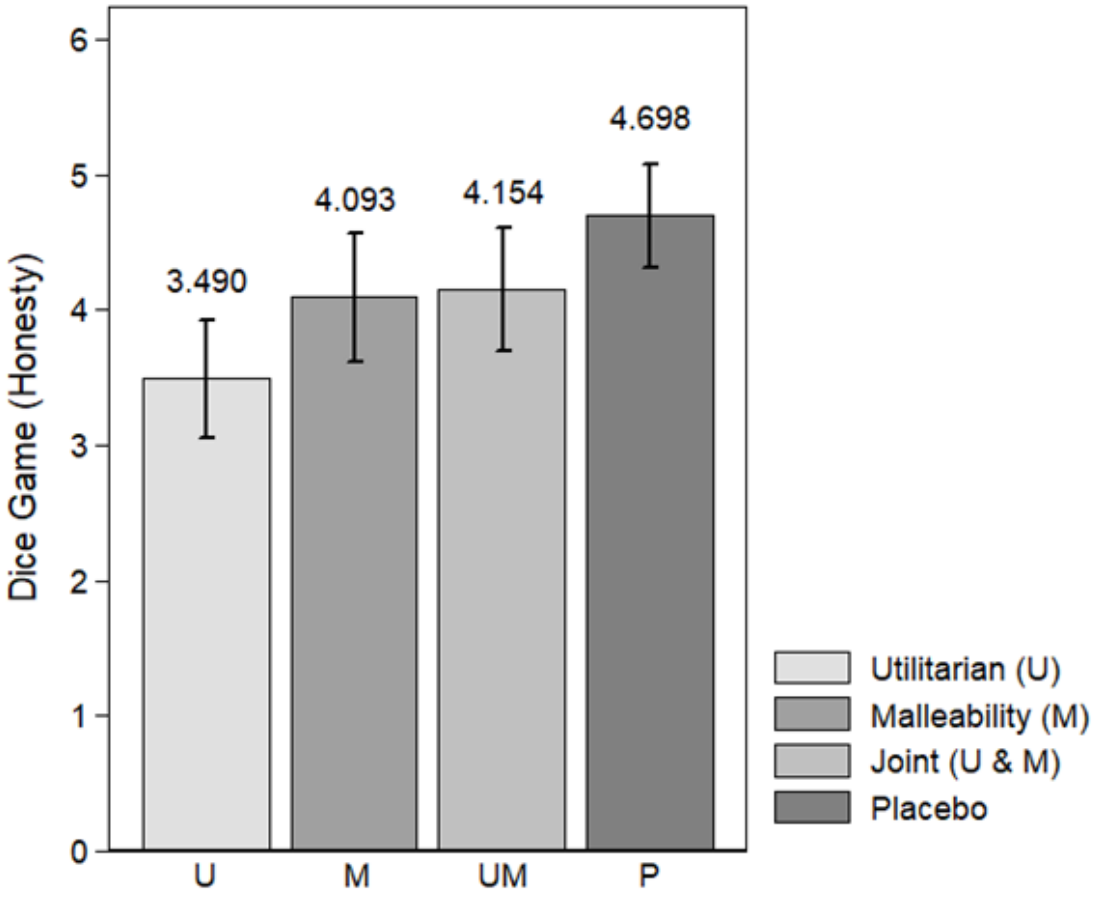


Panel B: Effect on fraction of "us versus them"



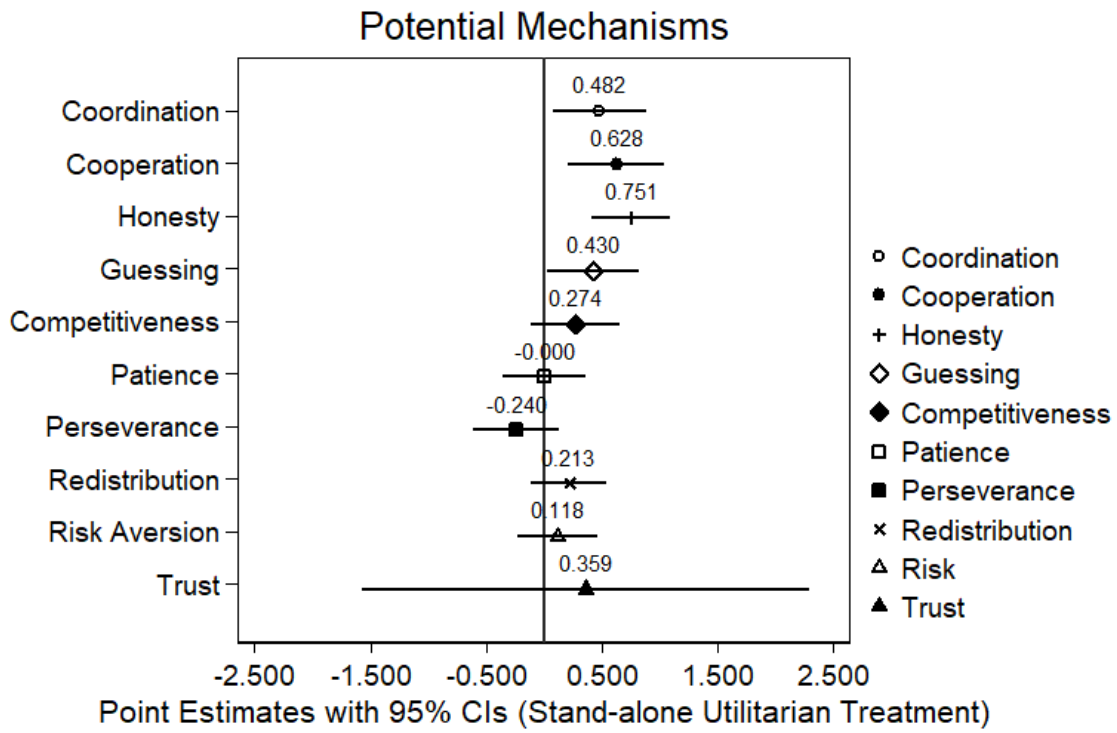
Note: All post treatment social media posts are considered up till 5 months following the interventions. Corresponding table-form representation of coefficient estimates with controls is presented in Appendix B.

Figure 5: **Effect on Dice Game**



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

Figure 6: **Exploration of Mechanisms**



Notes: The figure depicts the stand-alone utilitarian treatment effects and their 95% confidence intervals. Confidence intervals are based on standard errors clustered at the individual level (the unit of randomization). 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 – Dictator Games - Normalized

	Altruism Game		*Charity Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.064***	0.061***	0.178**	0.215**
	(0.023)	(0.021)	(0.088)	(0.091)
Stand-alone Malleability (*M*)	-0.020	-0.021	-0.011	-0.013
	(0.020)	(0.019)	(0.096)	(0.093)
Joint Treatment (*UM*)	-0.006	-0.018	-0.007	-0.046
	(0.010)	(0.012)	(0.096)	(0.093)
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

Robust standard errors clustered at individual level appear in 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 estimations obtained from OLS regressions includes 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: Impact of Treatments on Book Choice and Soft Skills Scores

	Empathy Book Choice		*Soft-Skills Scores*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.226**	0.232**	0.104***	0.116***
	(0.092)	(0.098)	(0.025)	(0.025)
Stand-alone Malleability (*M*)	0.030	0.040	0.005	0.003
	(0.096)	(0.097)	(0.026)	(0.025)
Joint Treatment (*UM*)	-0.017	-0.066	-0.004	-0.001
	(0.096)	(0.097)	(0.027)	(0.026)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
Mean of dep. var. (placebo)	0.460	0.460	0.541	0.541
p-value (test: *U* = *UM*)	0.009**	0.001**	0.000**	0.000**
p-value (test: *M* = *UM*)	0.622	0.264	0.755	0.882
p-value (test: *U* = *M*)	0.036**	0.042**	0.000**	0.000**
p-value (test: U*M* = *U* + *M*)	0.041**	0.013**	0.003**	0.002**

Robust standard errors clustered at individual level appear in 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 estimations obtained from OLS regressions includes 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: Results from the Field - Blood Donations, Orphanage Visits and Volunteering

Blood Donations

	Agreement to Donate		Appointment to Donate		Orphanage Visit	Volunteering in Schools
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.263***	0.062	0.284***	0.104	0.217**	0.226**
	(0.095)	(0.137)	(0.087)	(0.125)	(0.097)	(0.089)
Stand-alone Malleability (*M*)	0.081	0.063	0.041	0.062	0.003	0.104
	(0.086)	(0.129)	(0.077)	(0.127)	(0.091)	(0.086)
Joint Treatment (*UM*)	0.090	0.145	0.042	-0.026	0.052	0.091
	(0.087)	(0.127)	(0.075)	(0.105)	(0.090)	(0.085)
Blood Group Told (*T*)		-0.069		-0.059		
		(0.147)		(0.143)		
Blood Group Told X Stand-alone Utilitarian (*U X T*)		0.397**		0.355**		
		(0.192)		(0.173)		
Blood Group Told X Stand-alone Malleability (*M X T*)		0.040		-0.041		
		(0.183)		(0.169)		
Blood Group Told X Joint Treatment (*UM X T*)		-0.093		0.137		
		(0.175)		(0.153)		
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	205	205	205	205	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*

Robust standard errors clustered at individual level appear in brackets. The dependent variable in columns (1) and (2) are dummies that switch on for agreement to donate blood. The dependent variable 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 includes 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: Importance of Emotional Intelligence and Teamwork - Standardized

	Importance of Emotional Intelligence		*Teamwork Assessments*		*Quantitative Assessment*	
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.823***	0.840***	0.593***	0.615***	0.064	0.106
	(0.193)	(0.199)	(0.181)	(0.191)	(0.209)	(0.211)
Stand-alone Malleability (*M*)	0.189	0.159	-0.187	-0.197	-0.098	-0.078
	(0.204)	(0.217)	(0.180)	(0.191)	(0.189)	(0.190)
Joint Treatment (*UM*)	0.170	0.255	-0.334	-0.366	0.050	0.062
	(0.214)	(0.224)	(0.175)	(0.199)	(0.190)	(0.214)
Individual Controls	No	Yes	No	Yes	No	Yes
Observations	199	199	199	199	199	199
p-value (test: *U* = *UM*)	0.000***	0.002***	0.000***	0.000***	0.938	0.825
p-value (test: *M* = *UM*)	0.908	0.615	0.436	0.403	0.496	0.552
p-value (test: *U* = *M*)	0.000***	0.001***	0.000***	0.000***	0.453	0.403
p-value (test: U*M* = *U* + *M*)	0.001***	0.007***	0.007**	0.007**	0.764	0.906

Robust standard errors clustered at individual level appear in brackets. All dependent variables are standardized to mean 0 and standard deviation of 1. Dependent variable in Columns (1) and (2) is standardized variable to mean 0 and standard deviation 1 of the rating on a scale of 1 to 5 with 1 being not important at all and 5 as very important on the statement "How important do you think emotional intelligence i.e. the ability to monitor one's own and other people's emotions, to discriminate between different emotions is in public policy making?" *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. Dependent variable in Columns (3) and (4) present scores from regular public policy training courses at the Academy on the original scale of 0 to 10 on the workshop *Teams & Group Decisions*. This workshop simulates real decision these policymakers make in the field and assess the elite policymakers on their ability to respond as a team. Both teamwork are marked by a committee of senior bureaucrats and academics. Dependent variable in Columns (5) and (6) scores on *Quantitative Assessment* is reported. The assessment content included 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 includes 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: Mechanism - Impact of Treatments on Decision Making – Normalized

	Cooperation Game		*Coordination Game*		*Guessing Game*	
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.138***	0.136***	0.078**	0.065*	0.136**	0.116**
	(0.046)	(0.0489)	(0.033)	(0.035)	(0.062)	(0.058)
Stand-alone Malleability (*M*)	-0.042	-0.040	0.0213	0.018	0.040	0.037
	(0.040)	(0.040)	(0.029)	(0.031)	(0.054)	(0.055)
Joint Treatment (*UM*)	-0.003	-0.009	0.012	0.010	0.054	0.038
	(0.037)	(0.040)	(0.034)	(0.033)	(0.054)	(0.060)
Individual Controls	No	Yes	No	Yes	No	Yes
Observations	213	213	213	213	213	213
Mean of dep. var. (placebo)	0.535	0.535	0.849	0.849	0.085	0.085
p-value (test: *U* = *UM*)	0.001**	0.002**	0.045**	0.088	0.210	0.246
p-value (test: *M* = *UM*)	0.264	0.405	0.748	0.803	0.810	0.983
p-value (test: *U* = *M*)	0.000**	0.000**	0.048**	0.093	0.142	0.216
p-value (test: U*M* = *U* + *M*)	0.087	0.083	0.048**	0.117	0.151	0.173

Robust standard errors clustered at individual level appear in brackets. The dependent variables is 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 includes 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 7: Exploratory Analysis – Alternative Mechanisms – Normalized

	Competition Game	*Patience Game*	*Perseverance Game*	*Redistribution Game*	*Risk Aversion Game*	*Trust Game*
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.102	-0.002	-0.070	0.013	0.007	0.043
	(0.093)	(0.018)	(0.056)	(0.010)	(0.046)	(0.055)
Stand-alone Malleability (*M*)	0.014	-0.009	-0.057	0.009	-0.011	-0.026
	(0.880)	(0.022)	(0.060)	(0.009)	(0.052)	(0.058)
Joint Treatment (*UM*)	0.058	-0.014	0.025	0.008	-0.047	-0.015
	(0.536)	(0.019)	(0.070)	(0.008)	(0.053)	(0.053)
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

Robust standard errors clustered at individual level appear in brackets. The dependent variables is 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 includes 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:

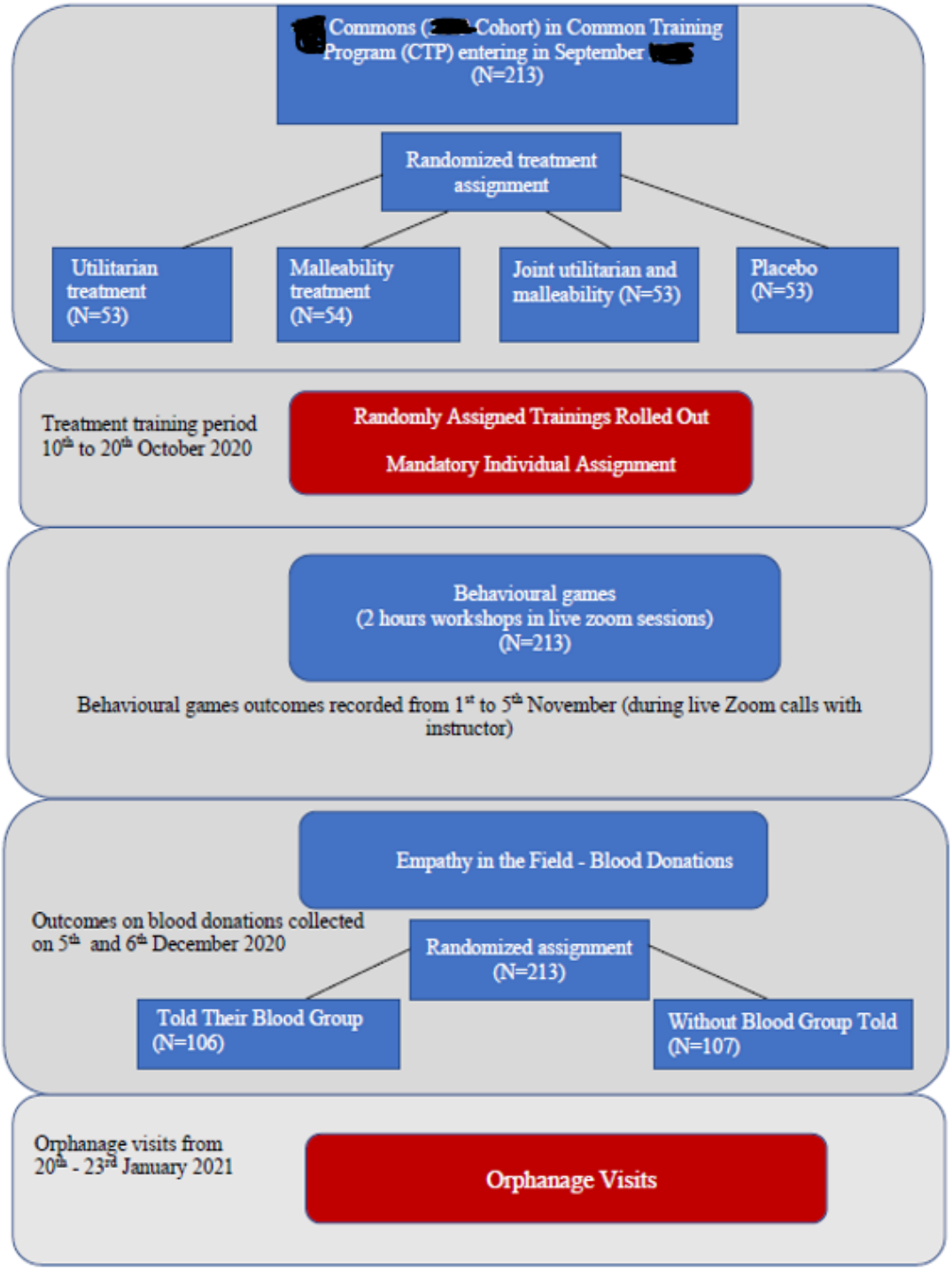
Training Effective Altruism

By Sultan Mehmood, Shaheen Naseer and Daniel Chen

Contents

B. Experimental Setup, Transcripts and Additional Tables

Table B1: Experimental Set-up



Note: The exact Commons Group that identifies the cohort and year of training is anonymized as per requests by the Federal Government of Pakistan and the Academy. It is available to the editor on request though a NDA or Non-Disclosure Agreement will need to be signed with Federal Government of Pakistan and FPSC.

Table B2: Transcript of Email sent by Director of Training Academy

Subject: Workshop - Mandatory Material

Dear Officers,

It is my pleasure to welcome you all to the upcoming [redacted] *workshop. With this email, I wanted to send you a link to a training lecture that you should watch very carefully and answer all accompanying questions before and after watching the lecture. Please note this is a mandatory individual training assignment so do NOT share the material or the accompanying questions/answers with anyone, especially your fellow officers. Failure to comply may lead to disciplinary action. I encourage you to watch the lecture twice so that all material contained in the lecture is well understood by you. Please click "finish" once you are completely done. The link with this training lecture is below: [link]*

Please access the link assigned to you by clicking on your name and entering your corresponding email. Good luck to you all!

Yours Sincerely,

Director [redacted]

Table B3: Utilitarian Treatment Transcript

I want to welcome all of you. I am your instructor for the soft skills workshop which we are starting next week. The purpose of sending you a presentation is to briefly walk you through some of the core concepts which will provide you the background knowledge that is compulsory for the upcoming workshop next week. And the first thing I want to do is, to make you comfortable. Although, this is compulsory lecture to get acquainted with the required material but there is nothing uptight about this presentation. I am really here for your benefit. I hope that is going to be a worthwhile experience for you. In this slide you see the topics that sort of headlines this presentation; We will talk about….What is empathy, Why it matters, why we need to talk about it. Then we will discuss qualitative or anecdotal evidence that is some examples from bureaucrats to underscore the importance of empathy. After presenting anecdotal evidence, we will discuss the empirical research on empathy. Ok to begin with: In modern economies the relevance of soft skills for organizational performance in the public and private sector is increasingly gaining traction. More than ever before, we are talking about organizational culture in a way that is not primarily focused on profits, regulations, processes and cognitive skills. To contexualize the discussion with some examples, let's take the example of some of the most profitable and biggest firms across the globe. In this table you see the names of companies across the globe which scored highest points in the empathy score. That means employees and employers in these firms are rated very high in empathy. Isn't it fascinating? "It is a puzzling question for economists why the most profitable and biggest firms rank so highly in empathy scores?" Why do firms who earn millions in profits also have high empathy? Is cut throatiness not going to get you more profits? Is the "rational self-interested notion of maximizing profit is most important? "Actually, it seems to be the case that soft skills are critical in all this!! "it may turn out that empathy boosts profit". This occurs because empathy equips stakeholders "employees and employers with the soft skills that allow the companies to navigate complex relationships and satisfy client needs and maintain employee trust and motivation". This empirical evidence is dispelling the view that it is being selfish and unemphatic to others is what will get you ahead in life. So, here are a few interesting definitions of empathy from different sources; this concept has been around for a while, and various religious beliefs teach us that it is something that we should practice as human beings towards others. There are different definitions of empathy in academic literature. Since there seems to be no universally agreed upon definition of empathy, "we don't need to go into the nitty gritty of each specific definition of empathy but in a nutshell empathy is putting yourself in another's shoes". It matters because the skill of empathy can help you succeed in your professional life. It can boost performance". That is to say, Empathy influences overall organizational performance and individual performance and well-being at a workplace. That is why, recent research is paying more and more attention to the effects of empathy on others. As we just saw in previous slide companies integrate empathy into their business strategies, because they think it'll help them to provide better services to their clients. We don't want to dwell too long on the private sector, but to bring it back to our context, of the importance of empathy for civil servants. Empathy is important for civil servants because public service organizations are challenging workplaces. That can be subject to emotionally demanding situations; you face demands of politicians, colleagues, clients etc. Empathy towards yourself, toward others, and

towards the citizens you serve can help you navigate this space better. It can help you at the job and it can improve services for your clients, because you consciously empathize with their needs, take their point of view, understand their concerns. This is especially relevant in a country where many people face severe hardship in daily lives and depend very much upon decisions you make!! We can find various examples of bureaucrats who are/were known for their empathic behavior towards others. For instance, Consider the example of Late KSD who recently passed away in the plane crash in Karachi? In his short career in the civil service he had made a name for himself as a "go getter" and person who delivered public service to the citizens. But not only Sherdil's repute was that of an honest, efficient, competent and above all always ready-to-help officer. He was famous for his empathy towards colleagues and citizens. Famous for helping his junior colleagues, going the extra mile when they were down and out. Here you have just one example where you have a high performing bureaucrat, admired by many for his devotion and performance, who is also well known for his empathy…Could it be that empathy and associated soft skills may have boosted his performance and helped him to deliver. It seems so. Systematic empirical research backs the idea that empathy can improve performance...also a related question is: why do private corporations train their employees in empathy? What is in for them? After all there is a Cutthroat competition in the corporate world for making profit. The point that I am trying to make it: Have you ever wondered why top multinational firms whose stated aim of existence is to maximize profits why are investing millions on "empathy" workshops? For example, at google, "Every new hire is trained in a "Google Empathy Lab". In the Google's empathy lab, employees are made to put on virtual reality googles and practice their perspective-taking or empathy. The employees are encouraged to take the perspective of homeless person and "see the world from the standpoint of the less fortunate". So in 21st century companies like Google may be investing in empathy to improve their profits and community engagement. So I made a rather bold empirical claim based on anecdotal accounts that empathy may be good for the companies financial performance. In fact a large body of research backs this up. For instance, in one prominent study at Stanford by Professor Zaki documents that empathy is more 'useful' than selfish behavior. It seems like a myth, being selfish is what will get you ahead but empathy and concern for other is key skill that those around you cherish. Empirical evidence shows that empathy benefits at different levels. First at a personal level, empathic people report to be much happier than less empathetic people. Second at a social level, empathic people have more fulfilling social lives than less empathetic people. Third, in fact empathic managers even have higher sales "Empathic managers are more productive and more successful " We have briefly touched on key findings from seminal studies on empathy that show the the very people who show empathy themselves are most successful. We would have time to go into detail of these study but I did want to give you a flavor of some research in this field...So we will go into detail of one of the studies. For instance, Scott et al., 2010 study 436 employees in a large US multinational. It was found more empathic managers had employees that reported to be happier but not only that they had "more sales"! Measuring higher on empathy scales rose sales up to 20%. Moreover, a sudden introduction of less empathic manager reduces work satisfaction, effort and sales. so the question is What is going on? (emphasis on it) Why is it that empathetic managers are not just report to be happier, their employees are happier but also they have higher sales! The answer is empathy generates trust and increase employee motivation and effort level. In a nutshell, bad bosses are bad for business. In addition, why high-stake decision makers like CEOs, elite bureaucrats and managers need empathy is that empathy is also a social good…Humans are "social animals". Empathy is social good which is valued by others If you are empathic, your subordinates will be more motivated to work with you FOR YOU! Empathy is mutually beneficial. Empathy helps you bring the best out of people. Only by taking the perspective of others can you realize the bottlenecks other people face in accomplishing their tasks and how they may overcome them. Mathieua and Babiak (2015) study 500 employees under 73 bosses. They found those bosses who scored high on "psychopathy test" had least productive employees and sales! "So, as you would have guessed nobody wants to work for bosses that are psychopathic or sociopathic". There are several studies that back the idea that if the team leader is empathic then the whole team performs better. Empathic leaders have better communication and trust with their employees and subordinates. Experimenters exposed empathic style of leadership and found employee quantity of hours put to work increased! Psychological research is suggesting that this may be due to "moral responsibility effect" (Fischer and Ravizza, 2000) It is relatively easier to shirk and "justify" your bad behavior with a bad boss not so easy with a good boss/ Another research on teams and performance, finds something very interesting. If you ask people on a team: who is the leader of the team? they are not likely to name the designated leader but the "effective leader who helped them out" in other words a colleague who was empathic to their needs, who may or may not be the designated leader. Again "humans are social animals", Empathy begets empathy. For you probationary officers this is of course not a surprise. You must have heard stories of the celebrated bureaucrats, the ones that made the difference! They incidentally also were revered not just their work ethnic and commitment to public service but also their empathy. 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.

Table B4: Malleability Treatment Transcript

I want to welcome all of you. I am your instructor for the soft skills workshop which we are starting next week. The purpose of sending you a presentation is to briefly walk you through some of the core concepts which will provide you the background knowledge that is compulsory for the upcoming workshop next week. And the first thing I want to do is, to make you comfortable. Although, this is compulsory lecture to get acquainted with the required material but there is nothing uptight about this presentation. I am really here for your benefit. I hope that is going to be a worthwhile experience for you. In this slide you see the topics that sort of headlines this presentation. We will talk about: What is empathy? Is empathy fixed? Before going in depth in the question of whether empathy is fixed in a person. I would mention some motivating examples that point towards the notion that empathy of person is not an immutable or unchangeable force of nature. After going through the anecdotal accounts, I will discuss some recent empirical research that shows whether empathy changes over time? We will specifically discuss Research on malleability of empathy . So, here are a few interesting definitions of empathy from different sources; this concept has been around for a while, various religious beliefs teach us that it is something that we should practice as human being toward others. There are different definitions of empathy in academic literature. Since there seems to be no universally agreed upon definition of empathy, "we don't need to go into nitty gritty of each specific definition of empathy but in a nutshell, empathy just means putting yourself in another shoes, its taking the perspective of others when making a decision". So the question is, Is empathy fixed? Throughout history anecdotal accounts suggest people can change, people can change in the level of empathy they show to others (From Religion: Hazrat Umar, Khalid bin Waleed (Islamic religious leaders) and their transformation from enemy of the Islam to the greatest champions of Islam. We can find various recent examples of people who are known for their drastic transformation; growing themselves into an empathetic personality. For instance, Consider the example of Majid Nawaz from being international terrorist to running the biggest counter-terrorism organization in Pakistan (Quilliem) that fights the battle against radicalization by presenting alternative narratives to radicalized youth and actual terrorists in jails across the world (see his book "The Radical" for his fascinating story). Many other examples across the world show that people can change in the level of empathy : for instance, some White supremacists in US becoming biggest fighters of minority rights. So, the question is what is going on? These example suggest that one can grow himself in empathy . So I made a rather bold empirical claim based on anecdotal accounts that empathy is not fixed. In fact a large body of research backs this up. For instance, in one prominent study at Stanford Zaki and co-authors show empathy is not fixed in a person. Several studies show empathy is nor fixed in a person (see e.g. Zaki and Ochsner, 2012). "Empathy is changeable and can be influenced over time. Empathy is not stable over one's lifetime. It can be developed and cultivated." Survey after surveys also show that empathy of populations changes over time. An important point is: Empathy doesn't come naturally in all situation: For instance: Sometimes we struggle with showing empathy for someone or considering their perspective. That's OK, empathy can be changed. If we don't feel empathy naturally, it doesn't mean that we are incapable of feeling it. empathy is changeable, and that understanding that it can sometimes be difficult to feel empathetic unless we work on it: is important step to developing this important life skill. Another important point is "Empathy is not a constant of nature determined by your upbringing alone, it rises and falls based on the environment around you"). For instance, in United States where most amount of data is available empathy scores have been falling for the last 30 years i.e. empathy in US now is about 50% of what it was 30 years ago. Why is it falling, if it is fixed? And it is not just one measure of empathy but all measures seem to follow this downward trend. This data convinced many psychologists that empathy is malleable, people can grow or fall in empathy. That is exactly what this graph indicates: that empathy is falling over time! If empathy is fixed theory is correct, this graph would not be downward trending. It should be a straight line. Essentially, this is inconsistent with the fixed empathy theory where empathy of individual and populations are fixed over time. This observed decline has put out of business all the psychological theories that had argued earlier empathy was fixed. We have briefly touched on key findings from the seminal study on empathy that show empathy is not fixed. I do want to give you some more flavor of cutting-edge research in this field. So, we will go into detail of couple of the studies. For instance: does empathy change? Empathy changed when they were given perspective of others (VR glasses, research article: Bernard et al., 2018). In the first study when

researchers gave virtual reality goggles to people and made them take perspective of others (e.g. see the lives through the eyes of homeless people and beggars), the level of empathy they showed to others skyrocketed both in surveys as well as high-stake decisions). Therefore, being open minded and willing to change and learn, is essential to grow in empathy and develop this skill. A seminal study from Stanford University shows that people who are most rigid in their believe that empathy cannot change in them or others are the least empathetic to begin with. People who believe empathy is inherent and unchangeable disengage from situations where empathy is difficult for them to experience. By contrast, people who believe empathy can be developed, they feel less threatened by perceiving that their empathic abilities are being challenged in a difficult situation. Another study shows that "Resilience training" increased empathy among radicalized Moroccan youth (research article: Feddes et al., 2015). This suggests that "People really change? it hints towards the notion that we need to revise this notion empathy cannot be changed and is fixed, the level of empathy an individual has is not destiny. This also suggests the answer of the puzzling question why the most profitable and biggest firms engage in empathy workshops and "waste" millions if empathy is unchangeable? Can it be that companies like Google and Facebook think empathy is malleable in people? They can inculcate these skills. So, coming back to the basic question we began with, can empathy evolve in a person? Commonsense stories, qualitative and quantitative evidence all point to one conclusion that empathy is malleable and it can change. Empathy is a skill that can be developed. Like any skill it needs work, to understand the needs of others and not just to best serve them but bring the best out of your subordinates. Learning "The art of empathy" needs practice. Qualitative and quantitative evidence backs the idea that empathy is not fixed but malleable. It is a skill that can be developed.

Table B5: Joint Utilitarian and Malleability Treatment Transcript

I want to welcome all of you. I am your instructor for the soft skills workshop which we are starting next week. The purpose of sending you a presentation is to briefly walk you through some of the core concepts which will provide you the background knowledge that is compulsory for the upcoming workshop next week. And the first thing I want to do is, to make you comfortable. Although, this is compulsory lecture to get acquainted with the required material but there is nothing uptight about this presentation. I am really here for your benefit. I hope that is going to be a worthwhile experience for you. In this slide you see the topics that sort of headlines this presentation; We will talk about, what is empathy, why it matters, why we need to talk about it. Then we will discuss qualitative and quantitative evidence to underscore the significance of empathy for your performance. In the last part of the presentation, I will discuss some recent empirical research that show whether empathy is beneficial for you and if empathy changes over time? So, here are a few interesting definitions of empathy from different sources; this concept has been around for a while, various religious beliefs teach us that it is something that we should practice as human being towards others. There are different definitions of empathy in academic literature. Since there seems to be no universally agreed upon definition of empathy, we don't need to go into nitty gritty of each specific definition of empathy but in a nutshell empathy is putting yourself in another shoes. It matters because the skill of empathy can help you succeed in your professional life. It can boost performance. That is to say, empathy influences overall organizational performance and individual performance and well-being at a workplace. That is why, recent research is paying more and more attention to the effects of empathy on others Empathy is important for civil servants because public service organizations are challenging workplaces. That can be subject to emotionally demanding situations; you face demands of politicians, colleagues, clients etc. Empathy towards yourself, toward others, and towards the citizens you serve can help you navigate this space better, it can help you at the job and it can improve services for your clients, because you consciously empathize with their needs, take their point of view, understand their concerns. This is especially relevant in a country where many people face severe hardship in daily lives and depend very much upon decisions you make. We can find various examples of bureaucrats who are/were known for their empathic behavior towards others. For instance, consider the example of Late KSD who recently passed away in the plane crash in Karachi. In his short career in the civil service he had made a name for himself as a "go getter" and person who delivered public service to the citizens. But not only Sherdil's repute was that of an honest, efficient, competent and above all always ready-to-help officer. He was famous for his empathy towards colleagues and citizens. Famous for helping his junior colleagues, going extra mile when they were down and out. Here

you have just one example where you have a high performing bureaucrat, admired by many for his devotion and performance, who is also well known for his empathy ...Could it be that empathy and associated soft skills may have boosted his performance and helped him to deliver. It seems so. Systematic empirical research backs the idea that empathy can improve performance. Also, a related question is: why do private corporations train their employees in empathy ? What is in for them? After all there is a cutthroat competition in the corporate world for making profit. The point that I am trying to make it: Have you ever wondered why top multinational firms whose stated aim of existence is to maximize profits why are investing millions on empathy workshops? •For example, at google, "Every new hire is trained in a "Google Empathy Lab". in the Google's empathy lab, employees are made to put on virtual reality googles and practice their perspective-taking or empathy. The employees are encouraged to take the perspective of homeless person and "see the world from the standpoint of the less fortunate. So in 21st century companies like Google may be investing in empathy to improve their profits and community engagement. So I made a rather bold empirical claim based on anecdotal accounts that empathy may be good for the company's financial performance. In fact a large body of research backs this up. For instance, in one prominent study at Stanford. Empathy benefits all involved. Professor Zaki documents that empathy is more 'useful' than selfish behavior. It seems like a myth, being selfish is what will get you ahead but empathy and concern for other is key skill that those around you cherish. Empirical evidence shows that Empathy benefits at different levels. First at a personal level, empathic ate people report to be much happier than less empathetic people. Second at a social level, empathic people have more fulfilling social lives than less empathetic people.Third, In fact empathic managers even have higher sales. Empathic managers are more productive and more successful. We have briefly touched on key findings from seminal studies on empathy that show empathy benefits the very people who show empathy themselves. We would have time to go into detail of these study but I did want to give you a flavor of some cutting edge research in this field...So we will go into detail of one of the studies. For instance, empathy is beneficial to all stakeholders– An example of a Research Study: Scott et al., 2010 study 436 employees in a large US multinational. It was found more empathic managers had employees that reported to be happier but not only that they had "more sales"! Managers measuring higher on empathy scales had sales up to 20% higher. Moreover, a sudden introduction of less empathic manager reduces work satisfaction, effort and sales. So, the question is what is going on? (emphasis on it) Why is it that empathetic managers are not just report to be happier, their employees are happier but also they have higher sales. The answer is empathy generates trust and increases employee motivation and level of effort. In a nutshell, bad bosses are bad for business. In addition, why high-stake decision makers like CEOs, elite bureaucrats and managers need empathy is that empathy is also a social good. Empathy is a "social good". Humans are "social animals". Empathy is social good which is valued by others. If you are empathic, your subordinates will be more motivated to work with you, for you! Empathy is mutually beneficial. Empathy helps you bring the best out of people. Only by taking the perspective of others can you realize the bottlenecks other people face in accomplishing their tasks and how they may overcome them. Mathieua and Babiak (2015) study 500 employees under 73 bosses. They found those bosses who scored high on "psychopathy test" had least productive employees and sales!•" So, as you would have guessed nobody wants to work for bosses that are psychopathic or sociopathic". Empathy reduces shirking by subordinates. There are several studies that back the idea that if the team leader is empathic then the whole team performs better. Empathic leaders have better communication and trust with their employees and subordinates. Experimenters exposed empathic style of leadership and found employee quantity of hours put to work increased! Psychological research is suggesting that this may be due to "moral responsibility effect"(Fischer and Ravizza, 2000). It is relatively easier to shirk and "justify" your bad behavior with a bad boss not so easy with a good boss. Another research on teams and performance, finds something very interesting. If you ask people on a team, who is the leader of the team? they are not likely to name the designated leader but the "effective leader who helped them out" in other words a colleague who was empathic to their needs, who may or may not be the designated leader. Again "humans are social animals, empathy begets empathy. Ok, so empathy may be useful but does it matter if empathy is fixed and determined force of nature? So the next question is whether is empathy fixed? Throughout history we have several examples that people can change, people can change in the level of empathy they show toward others (From Religion: Hazrat Umar, Khalid bin Waleed (Islamic religious leaders) and their transformation from enemy of the Islam to the greatest champions of Islam. We can find various recent examples of people who are known for their drastic transformation; growing themselves into an empathetic personality. For instance, Consider the example of Majid Nawaz from being international terrorist to running the biggest counter-terrorism organization in Pakistan (Quilliem) that fights the battle against radicalization by presenting alternative narratives to radicalized youth and actual terrorists in jails across the world (see his book "The Radical" for his fascinating story). Many other examples across the world show that people can change in the level of empathy : for instance, some White supremacists in US becoming biggest fighters of minority rights. So, the question is what is going on? These example suggest that one can grow himself in empathy. So I made a rather bold empirical claim based on anecdotal accounts

that empathy is not fixed. In fact a large body of research backs this up. For instance, in one prominent study at Stanford Zaki and co-authors show empathy is not fixed in a person. Several studies show empathy is nor fixed in a person (see e.g. Zaki and Ochsner, 2012). "Empathy is changeable and can be influenced over time. Empathy is not stable over one's lifetime. It can be developed and cultivated." Survey after surveys also show that empathy of populations changes over time. An important point is: Empathy doesn't come naturally in all situation: For instance: Sometimes we struggle with showing empathy for someone or considering their perspective. That's ok, empathy can be changed. If we don't feel empathy naturally, it doesn't mean that we are incapable of feeling it. empathy is changeable, and that understanding that it can sometimes be difficult to feel empathetic unless we work on it: is important step to developing this important life skill. Another important point is "Empathy is not a constant of nature determined by your upbringing alone, it rises and falls based on the environment around you"). For instance, in United States where most amount of data is available empathy scores have been falling for the last 30 years i.e. empathy in US now is about 50% of what it was 30 years ago. Why is it falling, if it is fixed? And it is not just one measure of empathy but all measures seem to follow this downward trend. This data convinced many psychologists that empathy is malleable, people can grow in empathy or they can fall in empathy. That is exactly what this graph indicates: that empathy is falling over time! If empathy is fixed theory is correct, this graph would not be downward trending. It should be a straight line. Essentially, this is inconsistent with the fixed empathy theory where empathy of individual and populations are fixed over time. This observed decline has put out of business all the psychological theories that had argued earlier empathy was fixed. We have briefly touched on key findings from the seminal study on empathy that show empathy is not fixed. I do want to give you some more flavor of cutting-edge research in this field. So we will go into detail of couple of the studies. For instance: does empathy change? empathy changed when they were given perspective of others (VR glasses, research article: Bernard et al., 2018). In the first study when researchers gave virtual reality goggles to people and made them take perspective of others (e.g. see the lives through the eyes of homeless people and beggars), the level of empathy they showed to others skyrocketed both in surveys as well as high-stake decisions). Therefore, being open minded and willing to change and learn, is essential to grow in empathy and develop this skill. A seminal study from Stanford University shows that people who are most rigid in their believe that empathy cannot change in them or others are the least empathetic to begin with. People who believe empathy is inherent and unchangeable disengage from situations where empathy is difficult for them to experience. By contrast, people who believe empathy can be developed, they feel less threatened by perceiving that their empathic abilities are being challenged in a difficult situation. Another study shows that "Resilience training" increased empathy among radicalized Moroccan youth (research article: Feddes et al., 2015). This suggests that "People really change? it hints towards the notion that we need to revise this notion empathy cannot be changed and is fixed, the level of empathy an individual has is not destiny. This also suggests the answer of the puzzling question why the most profitable and biggest firms engage in empathy workshops and "waste" millions if empathy is unchangeable? Can it be that companies like Google and Facebook think empathy is malleable in people? They can inculcate these skills. So, coming back to the basic question we began with, can empathy evolve in a person and it useful for you? Qualitative and quantitative evidence backs the idea that empathy is not fixed but malleable. It is a skill that can be developed. Qualitative and quantitative evidence also 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.

Table B6: Attrition in Blood Donation Responses

	Drop-Outs (not answering calls for blood donations)		
	(1)	(2)	(3)
Stand-alone Utilitarian (*U*)	0.019		0.011
	(0.033)		(0.057)
Stand-alone Malleability (*M*)	-0.006		0.004
	(0.024)		(0.060)
Joint Treatment (*UM*)	0.001		-0.003
	(0.024)		(0.051)
Blood Group Told			-0.030
			(0.048)
Blood Group Told X Stand-alone Utilitarian (*U X T*)		0.001	0.017
		(0.043)	(0.076)
Blood Group Told X Stand-alone Malleability (*M X T*)		-0.042*	-0.019
		(0.023)	(0.078)
Blood Group Told X Joint Treatment (*UM X T*)		-0.020	0.009
		(0.016)	(0.059)
Individual Controls	Yes	Yes	Yes
Observations	213	213	213
Mean of dep. var. (placebo)	0.019	0.019	0.019
p-value (test: *U* = *UM*)	0.544	0.549	0.549
p-value (test: *M* = *UM*)	0.783	0.286	0.286
p-value (test: *U* = *M*)	0.471	0.361	0.361
p-value (test: U*M* = *U* + *M*)	0.737	0.603	0.603

Robust standard errors clustered at individual level appear in brackets. The dependent variable is a dummy for not answering phone call for blood donation. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatments. The estimations obtained from OLS regressions includes 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.

Table B7: Impact on Empathy - Standardized

	Altruism Game		*Charity Game*	
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.551***	0.522***	0.374**	0.452**
	(0.198)	(0.177)	(0.184)	(0.192)
Stand-alone Malleability (*M*)	-0.174	-0.183	-0.023	-0.027
	(0.169)	(0.160)	(0.201)	(0.196)
Joint Treatment (*UM*)	-0.052	-0.151	-0.015	-0.097
	(0.087)	(0.107)	(0.201)	(0.194)
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**	0.001**	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

Robust standard errors clustered at individual level appear in brackets. The dependent variable is standardized to have a mean of zero and 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 includes 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: Mechanism - Impact of Treatments on Decision Making - Standardized

	Cooperation Game		*Coordination Game*		*Guessing Game*	
	(1)	(2)	(3)	(4)	(5)	(6)
Stand-alone Utilitarian (*U*)	0.438**	0.374**	0.482**	0.405*	0.628***	0.616***
	(0.198)	(0.187)	(0.203)	(0.214)	(0.211)	(0.217)
Stand-alone Malleability (*M*)	0.129	0.118	0.132	0.108	-0.189	-0.180
	(0.174)	(0.177)	(0.181)	(0.190)	(0.183)	(0.182)
Joint Treatment (*UM*)	0.174	0.123	0.074	0.063	-0.013	-0.042
	(0.174)	(0.192)	(0.207)	(0.204)	(0.169)	(0.179)
Individual Controls	No	Yes	No	Yes	No	Yes
Observations	213	213	213	213	213	213
Mean of dep. var. (placebo)	-0.185	-0.185	-0.172	-0.172	-0.049	-0.049
p-value (test: *U* = *UM*)	0.210	0.246	0.045**	0.088	0.001**	0.002**
p-value (test: *M* = *UM*)	0.810	0.983	0.748	0.803	0.264	0.405
p-value (test: *U* = *M*)	0.142	0.216	0.048**	0.093	0.000**	0.000**
p-value (test: U*M* = *U* + *M*)	0.151	0.173	0.048**	0.117	0.087	0.083

Robust standard errors clustered at individual level appear in brackets. The dependent variable is standardized to have a mean of zero and 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 includes 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: 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*)	0.211	-0.022	-0.228	0.339	0.029	0.160
	(0.194)	(0.182)	(0.184)	(0.248)	(0.181)	(0.203)
Stand-alone Malleability (*M*)	0.030	-0.089	-0.187	0.243	-0.041	-0.097
	(0.196)	(0.226)	(0.196)	(0.228)	(0.202)	(0.215)
Joint Treatment (*UM*)	0.121	-0.143	0.082	0.207	-0.184	-0.054
	(0.196)	(0.194)	(0.228)	(0.198)	(0.202)	(0.196)
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: U*M* = *U* + *M*)	0.677	0.907	0.096	0.171	0.534	0.683

Robust standard errors clustered at individual level appear in brackets. The dependent variable is standardized to have a mean of zero and 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 includes 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. These results are also illustrated in Figure 5. *** p<0.01, ** p<0.05, * p<0.1.

Table B10: Impact on Social Media Feeds – Original and Standardized Units

	Fraction of We vs I	*Fraction of Us vs Them*	*Fraction of We vs I (std.)*	*Fraction of Us vs Them (std.)*
	(1)	(2)	(3)	(4)
Stand-alone Utilitarian (*U*)	0.210	0.437***	0.483	1.138***
	(0.160)	(0.133)	(0.369)	(0.345)
Stand-alone Malleability (*M*)	0.046	0.061	0.106	0.161
	(0.183)	(0.160)	(0.421)	(0.416)
Joint Treatment (*UM*)	-0.025	-0.083	-0.057	-0.215
	(0.156)	(0.151)	(0.358)	(0.393)
Individual Controls	Yes	Yes	Yes	Yes
Observations	68	53	68	53
p-value (test: *U* = *UM*)	0.178	0.0001***	0.178	0.0001***
p-value (test: *M* = *UM*)	0.713	0.449	0.713	0.449
p-value (test: *U* = *M*)	0.337	0.021**	0.337	0.021**
p-value (test: U*M* = *U* + *M*)	0.303	0.013**	0.303	0.013**

Robust standard errors clustered at individual level appear in brackets. Dependent variable in in odd numbered columns presents fraction of we versus I, while even number columns have fraction of us versus them as dependent variable. The last two columns present results of dependent variables that are standardized to mean 0 and standard deviation of 1. *U*, *M* and *UM* are dummy variables indicating randomly assigned Utilitarian, Malleability and Joint treatment training lectures. The estimations includes 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 Perceived Importance of Prosociality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Risk Tolerance	*Patience*	*Perseverance*	*Altruism*	*Trust in others*	*Preference for redistribution*	*Cooperation*	*Competition*
Stand-alone Utilitarian (*U*)	0.198	-0.0196	-0.0655	0.00513	0.0548	-0.132	0.00567	-0.0111
	(0.271)	(0.0742)	(0.164)	(0.178)	(0.275)	(0.173)	(0.150)	(0.232)
Stand-alone Malleability (M)	-0.248	-0.127*	-0.303*	-0.106	-0.348	-0.291*	-0.242*	0.100
	(0.266)	(0.0728)	(0.161)	(0.175)	(0.270)	(0.169)	(0.147)	(0.227)
Joint Treatment (UM)	-0.325	-0.0543	-0.110	0.0266	-0.0785	-0.212	-0.0220	-0.292
	(0.269)	(0.0737)	(0.163)	(0.177)	(0.273)	(0.171)	(0.149)	(0.230)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	199	199	199	199	199	199	199	199

Standard errors clustered at individual level appear in brackets. 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 as very important on different traits with the statement "*How important do you think the following traits? 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 normalized treatment effects obtained from the seemingly unrelated 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: Randomization Inference

	(1) *Altruism Game*	(2) *Charity Game*	(3) *Empathy Book Choice*	(4) *Soft-Skills Scores*	(5) *Agreement to Donate*	(6) *Appointment to Donate*	(7) *Orphanage Visit*
Stand-alone Utilitarian (*U*)	0.061	0.215	0.232	0.151***	0.225	0.247	0.217
	(0.003) ***	(0.019) **	(0.019) **	(0.000) ***	(0.019) **	(0.005) ***	(0.026) **
	{0.007} ***	{0.028} **	{0.004} ***	{0.000} ***	{0.016} **	{0.005} ***	{0.025} **
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	213	213	213	213	207	207	213
Mean of dep. var. (placebo)	0.498	0.604	0.46	0.509	0.216	0.176	0.264

p-value corresponding to clustered standard errors at individual level appear in parentheses, while p-value from permutation inference are in 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 B13: Robustness to Multiple Hypothesis Testing

	Altruism Game	Charity Game	Cooperation Game	Coordination Game	Guessing Game	Competition Game	Patience Game	Perseverance Game	Redistribution Game	Risk Aversion Game	Trust Game
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Stand-alone Utilitarian (*U*)	0.0608	0.215	0.136	0.0655	0.116	0.106	-0.00136	-0.0696	0.0130	0.00894	0.0620
p-value	(0.003)***	(0.019)**	(0.005)***	(0.059)*	(0.047)**	(0.265)	(0.939)	(0.215)	(0.207)	(0.848)	(0.814)
Sharpened q-value	[0.090]*	[0.145]	[0.090]*	[0.094]*	[0.097]*	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
FWER p-value	{0.004}***	{0.018}**	{0.001}***	{0.043}**	{0.054}*	{0.271}	{0.945}	{0.253}	{0.120}	{0.861}	{0.813}
Stand-alone Malleability (*M*)	-0.0213	-0.0127	-0.0398	0.0175	0.0368	0.0178	-0.00866	-0.0571	0.00993	-0.0107	-0.187
p-value	(0.255)	(0.892)	(0.323)	(0.568)	(0.505)	(0.851)	(0.694)	(0.340)	(0.264)	(0.837)	(0.504)
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	{0.300}	{0.887}	{0.340}	{0.581}	{0.533}	{0.851}	{0.654}	{0.339}	{0.226}	{0.831}	{0.466}
Joint Treatment (*UM*)	-0.0172	-0.0461	-0.00924	0.0101	0.0381	0.0524	-0.0134	0.0249	0.00825	-0.0470	-0.230
p-value	(0.167)	(0.619)	(0.815)	(0.759)	(0.523)	(0.582)	(0.475)	(0.721)	(0.287)	(0.378)	(0.365)
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	{0.409}	{0.613}	{0.827}	{0.754}	{0.525}	{0.586}	{0.494}	{0.682}	{0.322}	{0.357}	{0.378}
Sample Size	213	213	213	213	213	213	213	213	213	213	213

Notes: 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 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.