Journal of Economic Psychology 64 (2018) 89–104

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Trust behind bars: Measuring change in inmates' prosocial preferences

Mario A. Maggioni *, Domenico Rossignoli, Simona Beretta, Sara Balestri

CSCC and DISEIS, Università Cattolica del Sacro Cuore, Milano, Italy

ARTICLE INFO

Article history:
Received 2 May 2017
Received in revised form 15 December 2017
Accepted 18 December 2017
Available online 19 December 2017

Keywords:
Trust
Lab-in-the-Field Experiment
Trust Game
Dictator Game
Longitudinal study
Prisons

ABSTRACT

The paper presents the results of a Longitudinal Lab-in-the-Field Experiment implemented between September 2015 and July 2016 performed in two State Prisons in California (USA) to measure change in prosocial preferences. A subset of eligible inmates willing to undertake GRIP (Guiding Rage Into Power) program, were randomly assigned to it. The paper tests whether the participation to this program (used as a treatment in the experiments) affects prosocial preferences of participants, with specific reference to trust. The results of a Difference-in-Differences (DID) estimation procedure show that trust significantly increased in GRIP participants compared to the control group. This result is robust to alternative estimation techniques and to the inclusion of an endogenous behavioral measure of altruism.

1. Introduction

A long debate persists in Economics on whether the discipline, by focusing on decisions and actions of rational actors, should abide by the principle of methodological individualism by which "all explanations must run in terms of the actions and reactions of individuals" (Arrow, 1994, p. 1). Most economic models also assume stable exogenous preferences since "without these assumptions, one cannot infer the causal connection between changes in opportunity sets and predicted changes in choices" (Dasgupta, Gangadharan, Maitra, & Mani, 2017, p. 17). Either or both of assumptions have been repeatedly challenged throughout the decades – from Veblen (1899) to Tarde (1902); from Hirschman (1984) to Bowles (1998) in the attempt to consider individuals as embedded in a socio-cultural environment and influenced by their relational networks.

Some recent approaches (notably arising from Behavioral Economics) have stressed the effects of forces outside the individuals which may shape their decisions and actions away from individualistic maximization procedures. In particular, Hoff and Stiglitz (2016) distinguish 3 types of economic literature: "(1) standard economics with *rational actors*, who have stable, exogenous preferences; (2) behavioral economics with *quasi-rational actors*, who make seemingly inconsistent choices; and (3) behavioral economics with *enculturated actors*, who have endogenous preferences, perception, and cognition" (Hoff & Stiglitz, 2016, pp. 27–28, our italics). The literature records many Lab Experiments highlighting the existence of quasi-rational actors and providing evidence of priming and framing phenomena (based on influences in the moment of decision), whereas there are few Field and/or Lab-in-the-Field Experiments exploring decisions and actions of enculturated actors, measuring the potentially durable impact of experience and exposure phenomena on their behaviors. Experience and exposure require actors to be observed in real life situations, where decisions are influenced by social identities, peer-effects, world-views and narratives.

* Corresponding author.
E-mail address: mario.maggioni@unicatt.it (M.A. Maggioni).

https://doi.org/10.1016/j.joep.2017.12.003
0167-4870/

This paper aims at contributing to this strand of literature by testing whether the participation to an offender accountability program (GRIP[1]), run by the US-based NGO *Insight-Out*, significantly changes prosocial attitudes (trust and altruism[2]) of violent crime offenders detained in two Californian state prisons. Changing preferences enhancing prosocial behavior of long-term prisoners is particularly difficult since, as shown by the literature (Haney, 2003), time spent in prison negatively affects inmates ability of trusting and engaging in enduring relationships. Our empirical results are in line with such claim by showing that trust and altruism decreased in prisoners not attending the GRIP program (i.e. in our control groups); however, the decrease of the relevant indicators is not statistically significant.

As outlined in the program's website, "the 10 months GRIP program functions as a peer education model where experienced students co-facilitate the classes and mentor newer students. The program employs a methodology that is called 'normative culture' wherein the students cultivate intrinsic motivation by being actively involved in both setting and enforcing the standards and norms that are integral to the course".[3] Thus GRIP may well qualify as an ideal test-bed for assessing the effects of experience and exposure in changing social preferences.

The way Insight-Out administers the program particularly fits the requirements for an experimental setting. The NGO's limited capacity (until 2016 Insight-Out has been able to manage only one GRIP class of about 25 inmates per prison each year) allows for the implementation of a Lab-in-the-Field Experiment[4] involving 42 treated (inmates participant to GRIP) and 38 controls (inmates not attending the program), in two Californian State prisons. All 80 inmates were surveyed twice, the first time before the beginning of GRIP and the second time about ten months later (after the end of the program). The experimental protocol, designed to run a panel data Difference-in-Differences analysis, envisaged a set of behavioral games, devised to elicit prosocial behavior (and in particular trust and altruism) and a series of questions based on a set of validated psychological scales of forgiveness and self-forgiveness.[5]

The novelty of our contribution is threefold:

1. *Longitudinal analysis*. We implemented a longitudinal study in the framework of prosocial behavioral games, concerning the change of prosocial attitudes and preferences over time, with reference to trust and altruism, experienced by people exposed to a "rehab" program. While longitudinal studies – according to Chuang and Schechter (2015, pp. 153–154) have been extensively used to test the stability of risk and time preferences, changes in social preferences have been investigated by a handful of papers. In particular, to the best of our knowledge, only one paper (Lönnqvist, Verkasalo, Walkowitz, & Wichardt, 2015) specifically uses the trust game in a longitudinal analysis, though as an indicator of risk attitudes; while another paper (Brosig-Koch, Riechmann, & Weimann, 2017) uses a Dictator Game in a longitudinal analysis whose main focus is the learning process of a quasi-rational actor.[6]
2. *Experience* and *Exposure* rather than *Priming* and *Framing*. Our Lab-in-the-Field Experiment, designed to measure the changes in trust and altruism displayed by a group of Californian inmates, is specifically designed to test the effect of the exposure to a different environment (the GRIP program as opposed to the prisons general environment) and on the experience of a strong sense of collective identity.
3. *Non-student population*. While conventional Lab Experiments use university students as targeted population, in this paper, we applied a behavioral economics set of games within an environment (state prisons) rarely investigated by the behavioral economics literature to a non-standard adult population (prison inmates) in which, as described by Hardin (2002) and Haney (2003), distrust among prisoners is often encouraged by corrections officers and managers to strengthen their control capacity.

The paper is organized as follows: Section 2 outlines the main features of the GRIP program, highlighting its core targets; Section 3 presents the research design and experimental methods; Section 4 presents the methodology to estimate the Average effect of the Treatment on the Treated (ATT); Section 5 provides results and robustness checks, and Section 6 provides a discussion of the main findings and conclusions.

2. *Guiding Rage Into Power* (GRIP): outline of the rehabilitation program

Jacques Verduin has been running the GRIP (*Guiding Rage Into Power*) Program in San Quentin Prison (California, USA) since 2011,[7] the founder of Insight-Out, an NGO based in the San Francisco's area. GRIP originates from its founder's vision that

[1] GRIP is the acronym for *Guiding Rage Into Power* an offender accountability program which is currently administered in seven Californian prisons and involves around 500 inmates. See more details in Section 2.

[2] In a somehow similar experimental setting, Blattman, Jamison, & Sheridan (2017) investigates the "malleability" of a number of noncognitive skills and preferences in criminally engaged adults.

[3] http://www.insight-out.org/index.php/programs/grip-program.

[4] Following the taxonomy proposed by Harrison and List (2004), a Lab-Like Field or Lab-in-the-Field experiment involves participants drawn from the field and asks them "to perform laboratory tasks that are not part of their day-to-day environment" (Viceisza, 2016, p. 836).

[5] The forgiveness-related questions are not analyzed in this paper due to their different nature and scope. Therefore, we have postponed their illustration to a future work.

[6] Both papers are based on Lab Experiments performed on university students. See point 3 of this list.

[7] His experience with prison programs is much longer: in 1997, he founded the Insight Prison Project (IPP), pioneering innovative in-prison rehabilitation programs designed to create transformational change among prisoners at California's San Quentin State Prison.

the lack of relations is the main driver of violence and unlawful behavior[8] and is classified as an "offender accountability program", according to the California Department of Corrections and Rehabilitation (CDCR).[9] GRIP aims at providing inmates with the skills to undo and prevent violent behavior so to become "agents of change", i.e. "people with skills to defuse conflicts around them".[10] In particular, the program focuses on the origins of behaviors and habits that are conducive to crime, with the specific purpose of undoing "the characteristic destructive behavioral patterns (...) that lead to transgressions".[10] The program usually spans over an "academic year", roughly ten months long (between September and July) and develops through weekly/fortnightly lessons, each one focused on a specific topic, aiming at (1) stopping violent behavior, (2) cultivating mindfulness, (3) achieving emotional intelligence and (4) understanding victim impact. GRIP targets unobservable behavioral traits, possibly affecting other-regarding preferences and beliefs. For this reason, the program is particularly suited to the experimental analysis we devised: neither trust, nor altruism are explicitly "taught" during GRIP classes, hence we expect experimental outcomes not to be driven by inmates' lip service to the explicit program content.[11] However, both trust and altruism can be indirectly affected by the program and the implementation of behavioral games is especially devised to elicit prosocial attitudes by the observed choices in a given set of tasks.

GRIP classes are held through a variety of didactic methods, spanning from traditional frontal lessons, to group-works and intervention of external guests.[12] Great emphasis and efforts are spent on creating a strong group identification, so that GRIP participants realize and experience that they are not alone, but part of a community that is pursuing the same objectives.[13] Every cohort of GRIP participants is named "Tribe" followed by a number that consists of how many years all the men (new participants as well as co-facilitators) have been incarcerated in any type of correctional facility: from juvenile detention to county jails to state prisons.[14] The inmates are also asked to work on their own or in small groups on specific assignments, in order to keep track of their progresses in achieving the four above mentioned goals. At the end of the course, a Graduation ceremony is held, involving inmates, relatives and the prison's warden. During the Graduation ceremony inmates, or "students" as they are called throughout the program, receive their title and diploma of "peace maker".

The GRIP program integrates three principal modalities: *Instruction* functions as a means to teach the information that is crucial to the programs theoretical framework. *Process* refers to the various exercises employed to work with a deep layer of emotional material that must be acknowledged, expressed and integrated in order for insight and understanding to occur. *Practice* anchors the acquired insights into a durable behavior by spending time learning how to embody what has been learned.

As other programs informed by the principles of Socio-Emotional Learning (SEL) and Cognitive Behavioural Theory (CBT), GRIP makes use of symbols, language, schemes that belong to street if not gang culture to shape inmates social identity and behavior: the Tribe is in some way an inverted or "flipped" and therefore positive version of a gang, in which tight interpersonal bonds are instrumental to the diffusion of a culture of accountability and peace as opposed to irresponsibility and violence, enhancing the construction of a collective identity that in turn promotes pro-social behavior (Blattman et al., 2017; Feigenberg, Field, & Pande, 2010; Hoff & Stiglitz, 2016; Paulle, 2017).

3. Experimental design

Before dealing with the description of the experimental design, it is worthwhile stressing that this paper does not directly aim at testing the impact or effectiveness of the program in reaching its core targets; but rather uses behavioral economics games, within an experimental setting, to measure effects of the program on changing inmates' prosocial preferences and attitudes. In particular, we focus on generalized trust, for two reasons: firstly, it is widely acknowledged (Fukuyama, 1995; Grimalda & Mittone, 2011; Guiso, Sapienza, & Zingales, 2008; Knack & Keefer, 1997; Putnam, 1993; Warren, 1999) as one of the ingredients of well-functioning societies, efficient and growing economic systems and effective participative institutions; secondly, prosocial attitudes usually seriously decline in prison since distrust is often promoted in prisons as a mean to facilitate control, by breaking loyalty links and interpersonal relationships among inmates. Therefore, the time spent in prison affects inmates' ability of trusting and engage in prosocial relationship. Haney (2003, p. 44) put it sharply:

> Because many prisons are clearly dangerous places from which there is no exit or escape, prisoners learn quickly to become hypervigilant and ever-alert for signs of threat or personal risk. Because the stakes are high, and because there are people in their immediate environment poised to take advantage of weakness or exploit carelessness or inattention, *interpersonal distrust* and suspicion *often result* (Our italics).

[8] A summary of both vision and mission of Insight-Out can found on the NGO's website. See https://goo.gl/uMFgna for further details.
[9] An independent presentation of the programs features can be found in Paulle (2017).
[10] See http://insight-out.org/index.php/programs/grip-program.
[11] No reference to trust and altruism/generosity can be found in the final "Pledge", signed by each inmate taking part in GRIP.
[12] For a detailed presentation on how the GRIP program works, refer to: https://vimeo.com/63489782.
[13] A recent analysis of the effectiveness of a somehow similar program (called STYL) in changing a series of noncognitive skills and preferences of criminally engaged men in Liberia can be found in Blattman et al. (2017).
[14] In a given class of around 30 men, the total often climbs higher than 600 years. In this particular way each cohort obtains its unique group identification, its name, that inmates often use to recall for years after the of end of the program. See Paulle (2017) for more details.

3.1. Identification strategy

GRIP does not target "specific" kinds of inmates, being virtually open to any inmate who applies. Before starting each program, Insight-Out offers a program orientation day in each prison to illustrate the main features of GRIP to all the potentially interested inmates. After the presentation of GRIP, inmates are asked to signal their interest and take an interview aimed at identifying mental disabilities or other severely impairing conditions. The actual number of selected participants (usually around 25 per prison) depends on both the NGO capacity constraints and prison security rules. This feature of the enrollment process particularly fits the methodological requirements of a pre-/post-treatment research design. The treatment (i.e. the participation to the program) was randomly assigned to inmates that signaled their interest and passed the interview, up to the filling of all available places. The random assignment of the treatment has been implemented according to the following procedure: after the motivational open day, Insight-Out provided us with information about the inmates (identified by an anonymous code) interested in taking part into GRIP. The resulting pool of inmates was randomly assigned to either treatment or control by balancing for ethnic group, to comply with CDCR requirements for prison programs and with GRIP's aim.

Henceforth, inmates enrolled in GRIP will be referred to as the "Treated" group (T); inmates who asked to enter the program but were not enrolled because of capacity constraints will be referred to as the "Waiting list" group (W); since, if the program is repeated in the same prison, they will have a chance to attend the program in the subsequent years. Finally, inmates who attended the orientation day, but did not show up for the interview, as the "No interview" group (N).

The experiment has been actually conducted on 80 inmates of two Californian State Prisons (Avenal and Mule Creek),[15] operated by the CDCR, according to the above mentioned three sample of inmates (T, W, N).[16] Since the perspective of this research especially focuses on changes in prosocial preferences and attitudes, we devised a longitudinal study by administering the same questionnaires twice to each inmate: the first time in September 2015, before the start of the treatment (i.e. the GRIP program), the second time in July 2016 after the end of the program. The first survey measures the initial level of the parameters of interest; the second survey measures whether a significant change in the parameters occurred after the treatment.

3.2. Incentivized tasks: Trust Game and Dictator Game

In order to assess whether GRIP affects the prosocial preferences and attitudes of inmates, we devised an experimental setting that included two incentivized tasks,[17] a Dictator Game and a Trust Game, that were administered separately, after instructions had been read loud and clarification questions have been answered.[18]

In the *Dictator Game* (Kahneman, Knetsch, & Thaler, 1986) a Proponent is provided with an exogenous endowment (usually a fixed monetary amount), he/she is matched to an anonymous partner who has received no endowment, and his/her choice consists of how to split the endowment between himself and the partner. Within the game theory framework, assuming a Proponent with self-regarding preferences, the Dictator Game has a unique Nash equilibrium in which the Proponent maximizes his/her payoff by keeping all the endowment and sending 0 to the partner. Therefore, deviation from the *selfish* equilibrium solution in the Dictator Game are used to measure empathy, altruism and/or pure generosity (Camerer, 2003; Forsythe, Horowitz, Savin, & Sefton, 1994; Guala & Mittone, 2010). In the *Trust Game*, also known as Investment Game (Berg, Dickhaut, & McCabe, 1995; Camerer & Weigelt, 1988), a Proponent is provided with an exogenous endowment, and he/she is matched to an anonymous partner who has received no endowment. The Proponent's decision now concerns whether and how much of his/her endowment to send to the anonymous partner; the Proponent is also informed that the experimenter will multiply (triple) any amount sent. The Respondent, once has received the total transfer (the amount sent by the Proponent, duly multiplied) is then told to choose if, and how much of the total amount received, to send back to the Proponent. Therefore, the final payoff of the Proponent will be equal to the initial endowment, less the amount sent to the Respondent, plus the amount sent back by the Respondent to the Proponent. This game has a unique sub-game perfect Nash equilibrium in which the Proponent maximizes his/her payoff by keeping all the endowment and sending 0 to the partner: in fact, solving by backward induction, since a *selfish* Respondent has no reason to send back any strictly positive amount, the Proponent maximizes his/her payoff by keeping the entire initial endowment. Sending a positive share of the initial endowment to anonymous partners signals agents propensity to interact with unknown partners, providing a proxy for generalized trust (Berg et al., 1995; Camerer, 2003; Johnson & Mislin, 2011), that has been defined as "the deliberate willingness of a decision maker to making himself vulnerable to the actions of another party" (Sutter & Kocher, 2007, p. 365).

Due to the impossibility of dealing with monetary payoffs (according to CDRC rules) we devised an alternative form of reward. After extensive consultation with the staff of NGOs working within prisons in the USA, we decided to use dehydrated

[15] In these two prisons GRIP is the only available offender accountability program, thus, by definition, we do not have to worry about possible spillovers arising across different programs.

[16] Given the small size of the W group compared with the T group, in the paper we also include a wider control group, obtained as the sum of W + N groups as a robustness check. This procedure is justified by the absence of significant statistical differences between W and N with respect to covariates and outcome variables (see Table 2 for details).

[17] As already mentioned in Section 1, since this research is part of a broader project, the experiment also included a series of questions based on a set of validated psychological scales of forgiveness (Chiaramello, Sastre, & Mullet, 2008; Mullet, Girard, & Bakhshi, 2004), and self-forgiveness (Pelucchi, Paleari, Regalia, & Fincham, 2013) that are not analyzed int this paper.

[18] For details on the experimental procedure see Appendix E: Supplementary materials.

soups (henceforth: soups) as rewards.[19] Since the experiment deals with non-monetary incentives, we devised a control for the "use value" (`soup like`) and the "exchange value" (`soup value`) that inmates attach to soups.

3.3. Procedures

The experiment has been implemented in paper-and-pencil through visits to both prisons in September 2015 (before the start of the program) and in July 2016 (after GRIP Graduation took place). The experiment has been administered by six students/interviewers purposely recruited in a local Community College.[20] Interviewers were thus independent both from Insight-Out and CDCR, to minimize the risk of possible strategic choices on the inmates sides. In both prisons, the same procedure has been applied, as follows.

Inmates were gathered in a room, equipped with tables and chairs, and sat down at an adequate distance from each other. The interviewers read aloud the instructions of each game before administering it, making sure that everyone in the room understood it well.[21] Inmates were informed that only one of the games would have been randomly drawn through the toss of a fair plastic coin at the end of the experiment session and rewarded: in this way, each inmate had the incentive to maximize his outcome in both games.

Both behavioral games were played in an anonymous double blind setting. Inmates were randomly assigned a code; Insight-Out staff held records about the matching between individual names and codes, but could not access individual outcome data (i.e. games results); the research team could access individual outcome data, matched with anonymous codes, but could not access individual names.

In both the Dictator and Trust games inmates were told they would have been randomly matched with anonymous partners. To administer the payment of the payoffs, we devised the following protocol. Once an inmate submitted his paper sheets,[22] an interviewer asked him to pick one out of five "reward booklets", randomly taken from a pile. The booklets included the anonymous partner's choices to be matched the inmate's decision in the Trust Game. These booklets contained the outcome of the choices of University non-academic staff (janitors, wardens, cleaners, cooks, etc.), that were asked to play the Trust Game as Respondent in a previous experiment run by the research team. Therefore, all the choices included in the booklet were generated by real people, and the inmates were aware of this fact.[23] The inmate was then asked to toss a plastic coin to select the game (either Dictator or Trust) to be rewarded and his choice in the randomly selected game was matched with the partner's choice, if needed.[24] Finally, the resulting payoff was paid in soups. On average, inmates earned 5.7 dried soups (modal value 5); each inmate took around twenty-five minutes to complete the experimental session, administration of the rewards included.

4. Analysis of the treatment effect

4.1. Estimation technique

The experimental framework allowed to perform a Difference-in-Differences (DID) analysis in which the change registered by the T group was compared to the change recorded by the W group (in order to control for the self-selection related to the individual willingness a/o motivation to join the program) and, more generally, to the change of a wider "control group" obtained by pooling inmates belonging to N and W groups (N + W). The average effect of the treatment on the treated (ATT) is the effect of GRIP on the inmates taking part into the program.

DID allows to test the ATT in a pre-/post-treatment setting, by controlling for possible confounding factors, including fixed time-invariant individual characteristics. Formally, the effect of a treatment (Treat) on an outcome (Y) can be tested through the model:

$$Y_{it} = \alpha + \beta Treat_{it} + \gamma Post_{it} + \rho(Post * Treat)_{it} + \delta X_i + \epsilon_{it} \quad (1)$$

where the subscripts i and t respectively refer to prisoners and periods (surveys); α represent the constant term; *Post* is the time dummy, taking value 1 for observations belonging to the second survey and zero otherwise; *Treat* is the treatment dummy; X_i are individual inmates' characteristics; and ϵ_{it} is the usual error term, while β, γ, ρ and δ are the parameters to be estimated. The ATT effect is estimated by the coefficient ρ. Formally, being g the groups in our sample, namely treated (T), waiting (W), and the wider control group (N + W) and t the two surveys (1 and 2), then ρ is defined as follows[25]:

[19] Dehydrated soups are highly valuable in maximum security prisons as they allow inmates to have a meal in the relative privacy of their cell. Moreover there is anecdotal evidence that these items are stored and traded with other inmates, thus they can be thought as imperfect substitutes of money in prison. Recent academic research provides further support for our choice: "Inmates are so unhappy with the quality and quantity of prison food that they receive that they have begun relying on ramen noodles a cheap, durable food product as a form of money in the underground economy" (Gibson-Light, 2017).

[20] Gavilan College, Gilroy, CA.

[21] Before starting the experiment, inmates were asked to play some trial sessions in order to verify their actual understanding of the instructions.

[22] Including the forgiveness-related questions, not presented in this paper.

[23] Being constrained by a limited amount of time to perform the experiment (due to strict security rules within prisons) we decided to administer the Trust Game in a strategic form (rather than interactive). To provide payoffs to the prisoners we matched their choices with those of UCSC non-academic staff who were administered a Trustworthiness Game (again in strategic form) before the prisons experiment.

[24] Only the Trust Game needs information about the partner's choice.

[25] The same applies to the wider control group (N + W).

$$\rho = (E[Y_{igt}|g = T, t = 2] - E[Y_{igt}|g = T, t = 1]) - (E[Y_{igt}|g = W, t = 2] - E[Y_{igt}|g = W, t = 1]) \quad (2)$$

A DID can be estimated in a regression framework by creating dummy variables for g and t. In this way it is possible to estimate a model that fully takes into account a set of possible confounding factors.

4.2. Data and variables

Eq. (1) is applied to the analysis of both the outcome of the Dictator Game and the Trust Game. In the Dictator Game, the outcome (i.e. the number of soups sent to the Respondent) is transformed into an indicator of the relative endowment (within the range 0–1) that the inmate shares with the anonymous partner. Analogously, the Trust Game yields a discrete outcome expressed in number of soups (between 1 and 10) that has been re-scaled in relative terms, into the 0–1 range.

The benchmark model of our analysis includes control variables related to the following individual characteristics that could affect the inmate's attitude to trust other people:

- age declared by inmate as of July 2016 (`age`);
- a dummy variable, coded 1 if inmates was not involved in a stable relationship, i.e. whether he was not married, separated/divorced, engaged or widowed (`single`);
- a control for personal preferences for soups (`soup like`);
- a control for the value of soups as means of exchange, independently of individual tastes (`soup value`).

In an extended specification, we also included dummies for ethnic identity, to control for possible cross-ethnic heterogeneity that could affect the propensity to trust an anonymous partner.[26]

Table 1 provides the summary statistics of the individual characteristics of the treated and of the two control groups and N + W; Table 2 records mean comparison tests for all three samples. In particular, the last three columns show that no significant differences occur between Treated and control groups on the main covariates, with the sole exception of `soup like`.

Since the literature (e.g. Camerer, 2003; Johnson & Mislin, 2011) suggests that the outcome of a Trust Game can be actually driven by pure altruism rather than trust, in order to estimate the effect of pure altruism on trust, we provide an extended version of our benchmark model that also includes the outcome of the Dictator Game as measured in the first survey.

Finally, since the experiment is repeated twice, we take into account possible autocorrelation of the error term at the individual level. To tackle this issue, all models are estimated with robust standard errors, clustered at the inmate level.

5. Results and discussion

5.1. Main results

Fig. 1 illustrates the distributions of the inmates' choices in the Trust Game, by survey and treatment group. While the W control group in the upper and lower panel report statistically indistinguishable distributions across waves (the Kruskal-Wallis test cannot reject the null hypothesis of identical distributions, with a p-value of .256), the distribution for the Treated group is clearly different in wave 1 and 2 of the survey (according to a Kruskal-Wallis test, with a p-value .066. Almost identical results are obtained by using a Wilkoxon rank-sum test. Table 3 reports the main statistics for both Dictator and Trust Game by group and wave.

Table 4 reports the results of the benchmark models' estimations: model (1) shows that, after the completion of the GRIP Program, the Treated inmates significantly increased the fraction sent to the anonymous partner by 0.212, as shown by the ATT coefficient in Table 4 (corresponding to a 20 per cent increase of the initial amount, given that soups were only available as integer numbers). This finding is robust to the identification of alternative control groups, as shown in model (2): the size of the two coefficients is substantially comparable, since both can be converted into an average increase of about 2 soups.[27] The ATT coefficient is higher than the minimum detectable effect computed through a power analysis based on established values for both average and standard deviation (Johnson & Mislin, 2011), see Figs. D.1 and D.2 in the Appendix.

As far as covariates are concerned,[28] Trust is negatively associated with `age`, although the size of the coefficient is very small. Furthermore, `single` is highly significant and negatively associated with Trust, suggesting that the absence of involvement in a stable romantic relationship (either present or past, successful or unsuccessful) signals a less trustful attitude (in the Sutter & Kocher, 2007 sense). The benchmark model's estimations thus support the occurrence of a "trust-increasing" effect of GRIP in participant inmates.

[26] Ethnic group identities are self-reported by inmates by choosing among not mutually exclusive categories, hence all included in the model estimation, taken from the US Census official definition of ethnic and racial groups as mandated by the Office of Management and Budget's (OMB) 1997 standards. For further information see https://www.census.gov/quickfacts/meta/long_RHI225215.htm.

[27] In fact – since dehydrated soups (the goods used in the actual experiment) are indivisible – the two coefficients may be interpreted as identical.

[28] Table 4 does not report further alternative specifications that include also the payoff earned in the first survey: this variable is not significant (see Appendix B, Table B1), implying that the time interval between the two sessions is large enough to cancel out potential "memory effect" of the previous payoff.

Table 1
Summary of samples characteristics, at baseline.

Variable	Obs	Mean	Std. Dev.	Min	Max
T Group					
Age	42	43.86	9.03	22	59
Single	42	0.57	0.50	0	1
Soup like	41	5.44	3.08	1	10
Soup value	41	5.59	3.54	1	10
W Group					
Age	22	47.59	8.63	30	63
Single	22	0.59	0.50	0	1
Soup like	22	7.64	2.77	1	10
Soup value	22	6.32	2.95	1	10
N + W Group					
Age	38	47.00	9.42	26	63
Single	38	0.50	0.51	0	1
Soup like	38	7.26	2.83	1	10
Soup value	38	6.39	2.95	1	10

Table 2
T-tests, by treatment group at baseline.

	Sample means			Mean comparison T-tests		
Variable	T	W	N + W	T/W	T/N + W	W/N
Age	43.86	47.59	47.00	−3.734	−3.143	1.403
Single	0.57	0.59	0.50	−0.020	0.071	0.216
Soup like	5.44	7.64	7.26	−2.197***	−1.824***	0.886
Soup value	5.59	6.32	6.39	−0.733	−0.809	−0.182
Ethnic groups						
White	0.40	0.41	0.34	−0.004	0.063	0.159
Native American/Alaska	0.12	0.14	0.13	−0.017	−0.013	0.011
Hawaiian Native/Pacific	0.10	0.14	0.11	−0.041	−0.010	0.074
Asian	0.10	0.14	0.13	−0.041	−0.036	0.011
Black/African American	0.33	0.45	0.34	−0.121	−0.009	0.267*
Hispanic/Latino	0.48	0.45	0.53	0.022	−0.050	−0.170

* p < .10.
** p < .05.
*** p < .01.

Before moving to the extended specification of our model, we tested the effect of GRIP on the outcome of the Dictator game through a DID estimation. As models (3) and (4) in Table 4 show, the ATT coefficient is not significant in the benchmark model for the two control groups.[29] Thus, the extended specification of the DID model presented in Table 5 treats `altruism` as an endogenously determined individual attitude and includes it among the covariates at the value observed at the first survey.

Since the observed outcome of a Trust Game can be driven by multiple motivations, such as other-regarding preferences, and *beliefs* over the trustworthiness of the anonymous partner (Fehr, 2009; Sapienza, Toldra-Simats, & Zingales, 2013), we estimate the effect of pure altruism on trust (Rabin, 1993) by including the outcome of a Dictator Game. In this way, we devise an extended model in which an endogenously determined attitude of the participants is included as a covariate for trust.

In the extended specification, as shown in Table 5, models (1)–(4) include the same covariates as in Table 4 as well as the endogenous proxy for altruism; models (5)–(8) also include dummy variables for self-reported ethnic identities. For all specifications, the table reports the coefficients for both the subsample of motivated inmates (W) and the wider control group (N + W). All specifications are estimated both through standard OLS, as common practice in the field, and through GLM for the binomial family with a logit link function. The latter estimation technique has been implemented following the suggestions by Papke and Wooldridge (1996), for bounded dependent variables.[30] ATT, the Average Treatment effect on

[29] This result is robust for alternative model specifications, including an extended set of covariates to control for ethnicity, and no covariates at all. Furthermore, both the Wilckoxon rank-sum and Kruskal-Wallis tests fail to reject the null hypothesis that the distributions of altruism for Treated and Controls are different across surveys. The power analysis based on average and standard deviation as in Ben-Ner, Putterman, Kong, and Magan (2004) implemented for null effect in the Dictator Game shows a minimum detectable effect that is slightly larger that the estimated ATT coefficient, given our sample size. Thus we may not observe a statistically significant treatment effect due to low statistical power.

[30] In fact our dependent variable is the fraction of soups sent to the other person and is therefore constrained within the range 0–1. Some researchers addressed this issue by implementing a Tobit model, but Papke and Wooldridge (1996) showed that censored regression techniques do not apply for variables with infeasible values beyond the censoring point. Therefore, GLM models are included in the table as a robustness check.

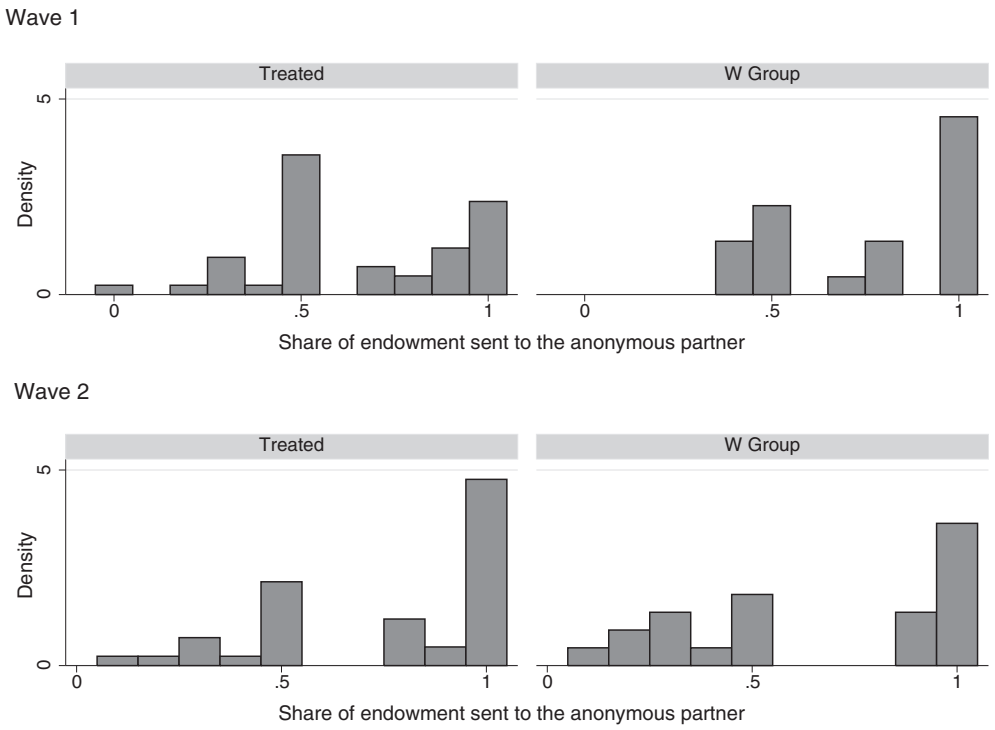

Fig. 1. Trust Game: distribution of inmates' choices.

Table 3
Summary statistics of Trust and Dictator.

	Wave 1			Wave 2		
Variable	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
			Treated group (T)			
Trust	42	6.55	2.75	42	7.60	2.85
Dictator	42	6.57	2.51	42	7.19	2.75
			Control group (W)			
Trust	22	7.64	2.50	22	6.59	3.38
Dictator	22	6.95	2.34	22	6.27	3.12
			Wider control group (N + W)			
Trust	38	7.42	2.52	38	6.66	3.16
Dictator	38	6.89	2.49	38	6.34	2.76

the Treated, is always positive and significant and it is robust to different model specifications. The coefficient of `altruism` is positive and significant, as expected.

In the extended version of the model reported on columns (5)–(8), none of the group dummies is significant, allowing to exclude effects of ethnic differences on trust.[31]

One may raise concerns on whether inmates – despite being told they were matched each with a real person, who had played, as a Respondent, the same game in a previously administered session, and whose answers were recorded in the "reward booklet" – were actually believing to interact with real persons, rather than simulated ones. Previous behavioral (Bottom, Holloway, Miller, Mislin, & Whitford, 2006) and neuroscience (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003) studies show that people behave differently according to their beliefs about the nature of partners (real vs. simulated persons). Empirical evidence shows that being convinced of playing against a machine or a simulated person

[31] Burks, Carpenter, and Verhoogen (2003, pp. 196–7) show that "Non-White participants exhibit less trust than whites in a mostly white environment". In our case, however, one should consider that the incarceration rate in the USA is higher for ethnic minorities. In an alternative model specification, not reported here, we run the same models as in columns (1)–(4) including a single ethnic dummy for Non-White inmate: the coefficient is always never statistically significant.

Table 4
Benchmark models' results: Trust and Dictator, Difference-in-Differences.

Dep. Var:	Trust		Dictator	
Control Group:	W	N + W	W	N + W
	(1)	(2)	(3)	(4)
ATT	0.212**	0.184**	0.132	0.119
	(0.087)	(0.075)	(0.090)	(0.073)
Age	−0.008***	−0.008***	−0.004	−0.004
	(0.003)	(0.002)	(0.003)	(0.003)
Single	−0.133**	−0.146***	−0.127**	−0.118**
	(0.059)	(0.049)	(0.053)	(0.047)
Prison dummy	−0.020	0.018	−0.150***	−0.088*
	(0.060)	(0.054)	(0.052)	(0.050)
Soup value	0.006	0.001	0.003	−0.006
	(0.009)	(0.008)	(0.008)	(0.008)
Observations	126	158	126	158
Adj. R-squared	0.08	0.11	0.09	0.07

Robust standard errors in parentheses, clustered at prisoners' level.
* p < .10.
** p < .05.
*** p < .01.

Table 5
Augmented Trust results: Difference-in-Differences.

Dep. var.: Trust	including altruism				including altruism & ethnic groups			
Control Group:	W		N + W		W		N + W	
Model:	OLS	GLM	OLS	GLM	OLS	GLM	OLS	GLM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATT	0.212**	0.199***	0.184**	0.171***	0.212**	0.201***	0.184**	0.172***
	(0.087)	(0.072)	(0.075)	(0.061)	(0.090)	(0.072)	(0.077)	(0.060)
Altruism	0.375***	0.414***	0.377***	0.407***	0.300**	0.343**	0.370***	0.395***
	(0.116)	(0.123)	(0.091)	(0.101)	(0.131)	(0.143)	(0.091)	(0.105)
Age	−0.006*	−0.006*	−0.006**	−0.007**	−0.006*	−0.007**	−0.006**	−0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Single	−0.122**	−0.127**	−0.138***	−0.142***	−0.140**	−0.142***	−0.143***	−0.140***
	(0.051)	(0.053)	(0.042)	(0.044)	(0.055)	(0.054)	(0.046)	(0.046)
Prison dummy	0.024	0.036	0.038	0.045	−0.013	−0.011	0.024	0.027
	(0.055)	(0.058)	(0.047)	(0.049)	(0.056)	(0.059)	(0.045)	(0.048)
Soup value	0.006	0.006	0.005	0.005	0.011	0.013	0.008	0.009
	(0.008)	(0.008)	(0.007)	(0.008)	(0.009)	(0.010)	(0.007)	(0.008)
White					0.071	0.069	0.043	0.046
					(0.062)	(0.065)	(0.052)	(0.059)
Native American/					−0.109	−0.095	−0.023	−0.028
Alaska					(0.099)	(0.103)	(0.074)	(0.071)
Hawaiian Native/					0.039	0.026	−0.009	−0.009
Pacific					(0.135)	(0.166)	(0.113)	(0.158)
Asian					−0.015	−0.033	−0.020	−0.043
					(0.137)	(0.133)	(0.087)	(0.105)
Black/					−0.079	−0.083	−0.079	−0.074
African American					(0.058)	(0.068)	(0.053)	(0.065)
Hispanic/					0.060	0.047	0.038	0.038
Latino					(0.065)	(0.071)	(0.050)	(0.058)
Observations	126	126	158	158	126	126	158	158
Adj. R-squared	0.17		0.21		0.18		0.21	
McFadden R-Sq.		0.29		0.12		0.31		0.13
AIC		1.009		0.974		1.086		1.037
BIC		−516.7		−695.2		−490.0		−666.8

Treatment and time dummies included. Clustered robust s.e. in parentheses.
OLS and GLM for the binomial family (link function: logit): for GLM models, the table reports marginal effects.
* p < .10.
** p < .05.
*** p < .01.

Table 6
Robustness check: Propensity Score Matching, dep. var.: trust.

Matching method:	NN† (1)	Radius† (2)	Kernel‡ (3)
ATT	0.188**	0.173**	0.109
	(0.094)	(0.082)	(0.077)
T-stat	2.00	2.10	1.42
Observations	79	79	79
Controls	20	29	37
Treated	41	32	41

Note: Common support always requested; balancing property satisfied.
Covariates: Altruism, age, single, soup value, ethnic group dummies.
Standard errors (†) or Bootstrap standard errors (‡) in parentheses.
* p < .10.
** p < .05.
*** p < .01.

exerts a downward bias on the fraction sent in the Trust Game (Johnson & Mislin, 2011, p. 873). However this does not hinder our results for two reasons: firstly, since our research design entails a DID, any potential downward bias occurring in both surveys is eliminated by the estimation technique; secondly, any residual downward effect would *a fortiori* strengthen our results.

5.2. Robustness checks

Another potential concern for the results shown in Table 5 is related to possible self-selection bias in the control group. As a robustness check we provide a Propensity Score Matching (PSM) procedure for the results. PSM is an estimation technique to evaluate the effectiveness of treatment in observational studies, extensively used since the seminal work by Rosenbaum and Rubin (1983). The aim is to estimate ATT by comparing treated and controls, conditioning on a set of relevant covariates (Becker & Ichino, 2002; Caliendo & Kopeinig, 2008). In order to proceed with the test we ignore the pre-/post- treatment design and consider only the outcomes of Trust in the second survey, by conditioning them on inmates' individual characteristics. The Propensity Score, i.e. the probability of being part of the Treated group, is therefore calculated by balancing the same covariates that we include in extended model specification (columns 5–8 in Table 5).[32] By imposing "common support" on Treated and Control, the actual matching will consider only Treated and Control with propensity scores within the range of the control group values. In this way, potential outliers in the Treated group are ignored, and the estimated ATT is robust to potential unobserved self-selection mechanisms.

Once the propensity score has been calculated[33] and the balancing property successfully tested, the choice of the matching algorithm is of pivotal importance since it affects both the bias and efficiency of the estimated ATT (Caliendo & Kopeinig, 2008). For this reason, Table 6 reports the estimated ATT according to three alternative matching algorithms extensively applied in the PSM literature, namely: 1-to-1 Nearest-Neighbor (NN), Radius Caliper and Kernel.[34]

Both models (1) and (2), shown in Table 6, report statistically significant ATT coefficients, very close to those estimated in Table 5 (ranging from 0.17 to 0.19) and rely on Nearest Neighbour and Caliper Radius algorithms that are based on similar procedures, to identify the most suitable observation in the control group for each treated one. These procedures are the most straightforward application of PSM and allow to compare similar individuals. For sake of completeness, column (3) in Table 6 reports the estimation of the ATT when the Kernel technique is applied. The lack of statistical significance in the Kernel model estimation is likely to be driven by the fact that this technique uses all the available information to generate the counterfactual outcome for the treated, thus including also potential "bad matches", i.e. controls with propensity scores very far from the treated (Caliendo & Kopeinig, 2008). Therefore, PSM overall confirms and strengthens our main result, supporting the evidence that trust, as measured by the fraction of soups sent to an anonymous partner in the Trust Game, increased in inmates participating to GRIP.

6. Conclusions

The present study tests whether GRIP, a specific offender accountability program for long-term sentenced inmates, implemented in two Californian State Prisons (Avenal and Mule Creek), changes the prosocial behavioral attitudes of

[32] Due to the limited size of the sample, the prison dummy has been not included since it would have further reduced the number of available propensity scores for the matching. All other covariates are indeed included in the propensity score estimation.
[33] The results of the logit estimation are shown in the Appendix. See Table C1.
[34] See Caliendo and Kopeinig (2008) for a survey of pros and cons of different matching algorithms.

inmates. The paper describes the results of a Lab-in-the-Field experiment, based on a longitudinal design spanning over a period of 10 months with a specific focus on altruism and trust. We chose to use the Trust Game since trust, commonly being acknowledged as an ingredient of the "cement of society" (Elster, 1989) and a lubricant of a social system (Arrow, 1974), it is hindered by long periods of incarceration (Haney, 2003), thus seriously questioning the rehabilitation function of the prison system. By focusing our analysis on trust, while controlling for altruism, we were able to effectively measure the effects of a 10 months exposure to the GRIP program in changing inmates' self-perceptions, social attitudes and preferences.

The research protocol envisaged administering twice a questionnaire including a set of behavioral situations ("games"), widely used in the experimental and behavioral economics literature, namely the Dictator Game (Kahneman et al., 1986) and the Trust Game (Berg et al., 1995), to 80 inmates, 42 treated (enrolled in the program) and 38 controls. The questionnaire was administered once before the start of the program; the second time at the end of the program.

A DID estimation procedure shows that trust significantly increased in GRIP participants compared to the control group. This result is robust to alternative estimation techniques and to the inclusion of an endogenous behavioral measure of altruism (measured by a Dictator Game).

On the one hand, the direct empirical results of the paper – based on a innovative longitudinal analysis conducted on a sample of non standard agents – support the claim that an offender accountability program, such as GRIP, produces beneficial effects on the inmates' prosocial preferences and attitudes, in addition and beyond its primary aims. On the other hand, the more general intuition that can be drawn from our analysis suggests that "exposure to a given social context shapes *who people are*" (Hoff & Stiglitz, 2016, p.26). Social contexts are the products and the repository of previous social, economic and psychological activities. "From this perspective, the situation, the context, or the environment then is not just an overlay on a set of basic psychological processes that provides the content for the processes. Instead (...) people think and feel and act in culture-specific (...) ways that are shaped by (...) meanings, practices, products, and institutions." (DiMaggio & Markus, 2010, p. 348). Therefore, programs like GRIP can be adopted as an effective instrument to or strengthen, in a relatively short span of time, prosocial preferences, attitudes and behaviors in inmates, while they are still in prison. Further they can also facilitate the rehabilitation process, in the long term, by fostering inmates' reintegration and re-socialization in their communities, thus potentially contributing to the reduction of recidivism.

Acknowledgments

The authors wish to thank C. Beretta, M. Colagrossi, E. Colombo, P.A. David, B. Paulle, V. Rotondi, L. Sullivan, F. Trombetta and the participants to the CSCC LBR Workshop held in Milan (February, 2017), as well as the editors and two anonymous referees of this journal for their comments and observations, which greatly helped improving the quality of the paper. The usual caveats apply. We thank E. Esposto for research assistantship. Financial support by the Fetzer Institute (Project # 3520.00) and the D3.2 Competitive Research Funds Università Cattolica del Sacro Cuore are gratefully acknowledged.

Appendix A. Correlation matrix

See Table A1.

Table A1
Pairwise correlation of main variables.

	Trust	Altruism	Age	Single	Soup like
Altruism	0.349				
	(0.000)				
Age	−0.235	−0.174			
	(0.003)	(0.030)			
Single	−0.234	−0.002	−0.114		
	(0.003)	(0.982)	(0.151)		
Soup like	0.027	−0.119	0.206	−0.118	
	(0.738)	(0.135)	(0.009)	(0.139)	
Soup value	0.076	−0.14	−0.005	−0.202	0.350
	(0.345)	(0.079)	(0.952)	(0.011)	(0.000)

P-values in parentheses.

Appendix B. Further robustness checks

See Tables B1–B3.

Table B1
Difference-in-Differences, same models as in Table 5, including wave 1 payoffs.

Dep. Var.: Trust	including altruism				including altruism & ethnic groups			
Control Group:	W		N + W		W		N + W	
Model:	OLS (3)	GLM (4)	OLS (1)	GLM (2)	OLS (7)	GLM (8)	OLS (5)	GLM (6)
ATT	0.212**	0.199***	0.184**	0.171***	0.212**	0.200***	0.184**	0.172***
	(0.088)	(0.071)	(0.076)	(0.060)	(0.090)	(0.070)	(0.077)	(0.060)
Altruism	0.349***	0.396***	0.373***	0.410***	0.270**	0.311**	0.366***	0.396***
	(0.119)	(0.125)	(0.090)	(0.101)	(0.131)	(0.144)	(0.089)	(0.103)
Payoff	0.008	0.010*	0.006	0.008	0.008	0.011*	0.006	0.007
	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
⋮								
Observations	126	126	158	158	126	126	158	158
Individuals		63		79		63		79
Adj. R-squared	0.19		0.22		0.24		0.25	
McFadden R-Squared		0.30		0.12		0.31		0.14
AIC		1.016		0.981		1.093		1.046
BIC		−512.9		−690.9		−486.2		−662.3

Treatment and time dummies included. Clustered robust s.e. in parentheses.
OLS and GLM for the binomial family (link function: logit): for GLM models, the table reports marginal effects.
* p < .10.
** p < .05.
*** p < .01.

Table B2
Comparison of ATT in different prisons, models as in Table 5.

				AVENAL			MULE CREEK		
	Control Group:	Model:		ATT	S.E	N	ATT	S.E.	N
Including altruism	W	(1)	OLS	0.227*	(0.127)	74	0.198*	(0.109)	52
		(2)	GLM	0.208**	(0.100)	74	0.181**	(0.082)	52
	N + W	(3)	OLS	0.175	(0.107)	98	0.230**	(0.101)	60
		(4)	GLM	0.158*	(0.084)	98	0.205***	(0.074)	60
Including altruism & ethnic groups	W	(5)	OLS	0.227*	(0.133)	74	0.198*	(0.116)	52
		(6)	GLM	0.198**	(0.098)	74	0.176**	(0.080)	52
	N + W	(7)	OLS	0.175	(0.111)	98	0.230**	(0.106)	60
		(8)	GLM	0.158*	(0.083)	98	0.199***	(0.072)	60

The table reports the estimated ATT in the two prisons taken separately.
The models are the same as in Table 5, including covariates (not reported here for reason of space), estimation techniques and clustered standard errors.
* p < .10.
** p < .05.
*** p < .01.

Table B3
'Extended' Trust results including time-varying (t.v.) dictator: Diff-in-Diff.

Dep. Var.: Trust	including t.v. dictator				including t.v. dictator & ethnic groups			
Control Group:	W		N + W		W		N + W	
Model:	OLS (1)	GLM (2)	OLS (3)	GLM (4)	OLS (5)	GLM (6)	OLS (7)	GLM (8)
ATT	0.123*	0.115*	0.104*	0.096*	0.105*	0.092*	0.129*	0.116*
	(0.066)	(0.061)	(0.058)	(0.053)	(0.060)	(0.055)	(0.068)	(0.062)
T.v. dictator	0.674***	0.727***	0.674***	0.742***	0.660***	0.731***	0.630***	0.689***
	(0.085)	(0.095)	(0.071)	(0.086)	(0.073)	(0.085)	(0.090)	(0.094)
Age	−0.005*	−0.006**	−0.006**	−0.006**	−0.005**	−0.006**	−0.005	−0.006*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Single	−0.047	−0.060	−0.066*	−0.077**	−0.067*	−0.070*	−0.056	−0.057
	(0.041)	(0.041)	(0.034)	(0.035)	(0.038)	(0.037)	(0.044)	(0.045)
Prison dummy	0.082	0.090	0.077*	0.086*	0.065	0.074	0.050	0.055
	(0.053)	(0.055)	(0.044)	(0.045)	(0.042)	(0.045)	(0.056)	(0.058)
Soup value	0.004	0.005	0.005	0.006	0.007	0.009	0.009	0.012
	(0.007)	(0.008)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.009)

Table B3 (*continued*)

Dep. Var.: Trust	including t.v. dictator				including t.v. dictator & ethnic groups			
Control Group:	W		N + W		W		N + W	
Model:	OLS (1)	GLM (2)	OLS (3)	GLM (4)	OLS (5)	GLM (6)	OLS (7)	GLM (8)
White					0.033	0.037	0.062	0.057
					(0.046)	(0.052)	(0.053)	(0.055)
Native American/					0.002	0.011	−0.077	−0.053
Alaska					(0.080)	(0.066)	(0.104)	(0.102)
Hawaiian Native/					−0.008	−0.054	0.049	0.015
Pacific					(0.099)	(0.153)	(0.111)	(0.150)
Asian					−0.041	−0.047	−0.067	−0.092
					(0.064)	(0.084)	(0.115)	(0.148)
Black/					−0.050	−0.035	−0.041	−0.036
African American					(0.045)	(0.056)	(0.048)	(0.060)
Hispanic/					0.042	0.055	0.070	0.078
Latino					(0.046)	(0.053)	(0.053)	(0.059)
Observations	126	126	158	158	158	158	126	126
Adj. R-Squared	0.43		0.46		0.47		0.44	
Mc-Fadden R-Sq.		0.39		0.23		0.24		0.39
AIC		0.896		0.862		0.930		0.979
BIC		−531		−712.9		−683.8		−503.5

Treatment and time dummies included. Clustered robust s.e. in parentheses.
OLS and GLM for the binomial family (link function: logit): for GLM models, the table reports marginal effects.
* p < .10.
** p < .05.
*** p < .01.

Appendix C. Propensity score estimation

See Table C1 and Fig. C.1.

Table C1
Propensity Score estimation: Logit model.

Dep. var.: Covariates	Treatment (1)
Altruism	−0.891
	(1.026)
Age	−0.055*
	(0.029)
Single	0.015
	(0.513)
Soup value	−0.107
	(0.080)
White	0.471
	(0.616)
Native American/Alaska	0.477
	(0.859)
Hawaiian Native/Pacific	0.224
	(1.575)
Asian	−1.225
	(1.366)
Black/African American	0.190
	(0.637)
Hispanic/Latino	−0.140
	(0.573)
Observations	79
Log-Likelihood	−51.24

Note: Logit estimation, marginal effects reported.
Standard errors in parentheses.
** p < .05.
*** p < .01.
* p < .10.

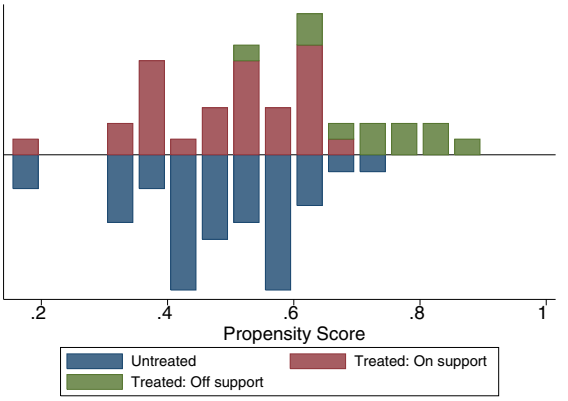

Fig. C.1. Estimated propensity scores: common support.

Appendix D. Power analysis

See Figs. D.1 and D.2.

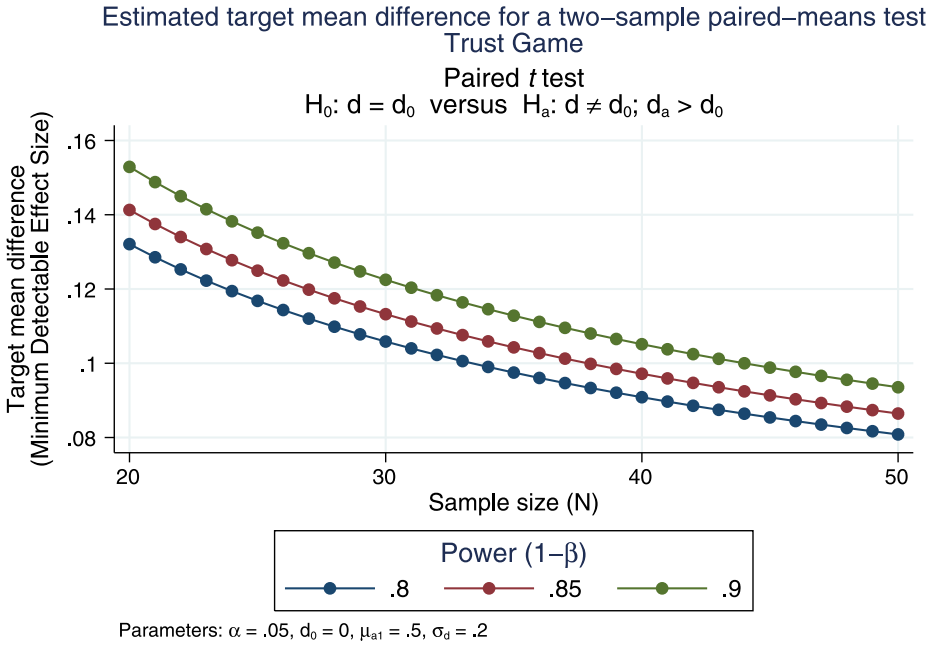

Fig. D.1. Power analysis: minimum detectable effect size of trust.

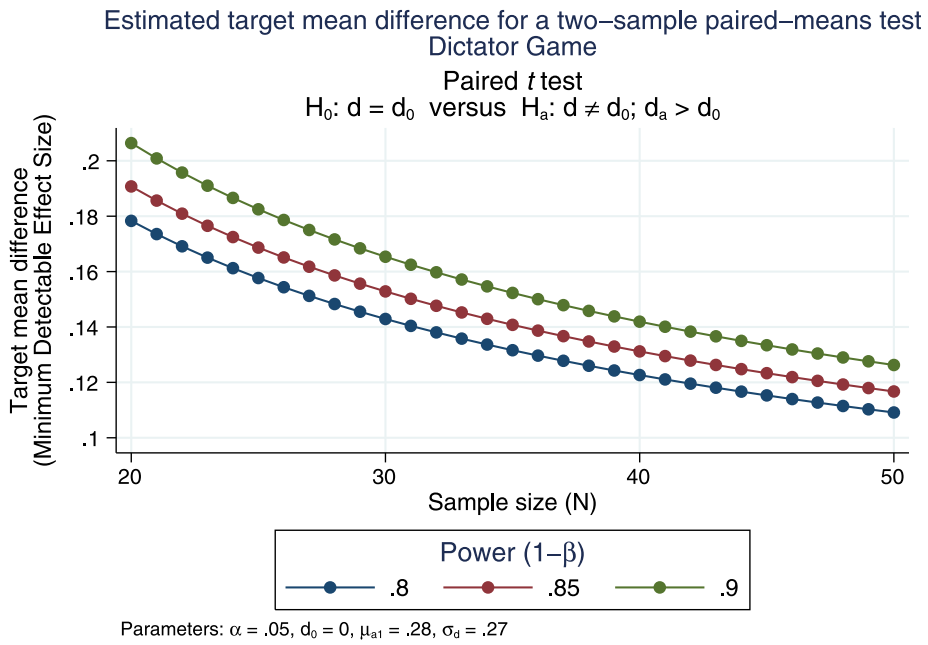

Fig. D.2. Power analysis: minimum detectable effect size of dictator.

Appendix E. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.joep.2017.12.003.

References

Arrow, K. J. (1974). *The limits of organization*. New York: WW Norton & Company.
Arrow, K. J. (1994). Methodological individualism and social knowledge. *The American Economic Review, 84*(2), 1–9.
Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal, 2*(4), 358–377.
Ben-Ner, A., Putterman, L., Kong, F., & Magan, D. (2004). Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization, 53*(3), 333–352.
Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122–142.
Blattman, C., Jamison, J. C., & Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia. *The American Economic Review, 107*(4), 1165–1206.
Bottom, W. P., Holloway, J., Miller, G. J., Mislin, A., & Whitford, A. (2006). Building a pathway to cooperation: Negotiation and social exchange between principal and agent. *Administrative Science Quarterly, 51*(1), 29–58.
Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature, 36*(1), 75–111.
Brosig-Koch, J., Riechmann, T., & Weimann, J. (2017). The dynamics of behavior in modified dictator games. *PloS One, 12*(4), e0176199.
Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization, 51*(2), 195–216.
Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.
Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.
Camerer, C., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica, 56*(1), 1–36.
Chiaramello, S., Sastre, M. T. M., & Mullet, E. (2008). Seeking forgiveness: Factor structure, and relationships with personality and forgivingness. *Personality and Individual Differences, 45*(5), 383–388.
Chuang, Y., & Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics, 117*, 151–170.
Dasgupta, U., Gangadharan, L., Maitra, P., & Mani, S. (2017). Searching for preference stability in a state dependent world. *Journal of Economic Psychology, 62*, 17–32.
DiMaggio, P., & Markus, H. R. (2010). Culture and social psychology: Converging perspectives. *Social Psychology Quarterly, 73*(4), 347–352.
Elster, J. (1989). *The cement of society: A survey of social order*. Cambridge, UK: Cambridge University Press.
Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association, 7*(2–3), 235–266.
Feigenberg, B., Field, E. M., & Pande, R. (2010). *Building social capital through microfinance*. NBER working paper 16018.
Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*(3), 347–369.
Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. New York: Free Press Paperbacks.
Gibson-Light, M. (2017). *Ramen politics: Informal money and logics of resistance in the contemporary american prison*. Qualitative Sociology.
Grimalda, G., & Mittone, L. (2011). Generalized trust: An experimental perspective. In *Social capital, corporate social responsibility, economic behaviour and performance* (pp. 260–281). London: Palgrave Mac-Millan.
Guala, F., & Mittone, L. (2010). Paradigmatic experiments: The dictator game. *The Journal of Socio-Economics, 39*(5), 578–584.
Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance, 63*(6), 2557–2600.
Haney, C. (2003). The psychological impact of incarceration: Implications for post-prison adjustment. *Prisoners Once Removed: The Impact of Incarceration and Reentry on Children, Families, and Communities, 33*, 66.
Hardin, R. (2002). *Trust and trustworthiness*. New York: Russell Sage Foundation.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.
Hirschman, A. O. (1984). Against parsimony: Three easy ways of complicating some categories of economic discourse. *Bulletin of the American Academy of Arts and Sciences, 37*(8), 11–28.
Hoff, K., & Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior & Organization, 126*, 25–57.
Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865–889.
Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review, 76*(4), 728–741.
Knack, S., & Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics, 112*(4), 1251–1288.
Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *Journal of Economic Behavior & Organization, 119*, 254–266.
Mullet, E., Girard, M., & Bakhshi, P. (2004). Conceptualizations of forgiveness. *European Psychologist, 9*(2), 78–86.
Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*(6), 619–632.
Paulle, B. (2017). Stumbling on the rehabilitation gold? Foucault vs. foucault in San Quentin and beyond. *Ethnography, 18*(4), 471–492.
Pelucchi, S., Paleari, F. G., Regalia, C., & Fincham, F. D. (2013). Self-forgiveness in romantic relationships: It matters to both of us. *Journal of Family Psychology, 27*(4), 541.
Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton: Princeton University Press.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review, 83*(5), 1281–1302.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755–1758.
Sapienza, P., Toldra-Simats, A., & Zingales, L. (2013). Understanding trust. *The Economic Journal, 123*(573), 1313–1332.
Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior, 59*(2), 364–382.
Tarde, G. (1902). *Psychologie conomique*. Paris: Flix Alcan, diteur.
Veblen, T. (1899). *The theory of the leisure class: An economic study in the evolution of institutions*. New York: Macmillan.
Viceisza, A. C. (2016). Creating a lab in the field: Economics experiments for policymaking. *Journal of Economic Surveys, 30*(5), 835–854.
Warren, M. E. (1999). *Democracy and trust*. Cambridge: Cambridge University Press.