# Rehabilitation and social behavior: Experiments in prison

Loukas Balafoutas[#]

*Department of Public Finance, University of Innsbruck, Austria*

Aurora García-Gallego

*LEE & Department of Economics, University Jaume I, Spain*

Nikolaos Georgantzis

*Burgundy School of Wine & Spirits Business, Dijon, France and LEE & Department of Economics, University Jaume I, Spain*

Tarek Jaber-Lopez

*Department of Public Finance, University of Innsbruck, Austria*

Evangelos Mitrokostas

*Department of Economics and Finance, University of Portsmouth, United Kingdom and The Hellenic Open University, Greece*

[#]Corresponding Author: *Universitätsstrasse 15, 6020 Innsbruck, Austria. Fax: +4351250770399. Tel.: +4351250770206. E-mail: Loukas.Balafoutas@uibk.ac.at.*

**Abstract**

Despite the economic and social significance of crime reduction and criminals' rehabilitation, research evaluating the effects of incarceration on behavior is surprisingly scarce. We conduct an experiment with 105 prison inmates and complement it with administrative data in order to explore several aspects of their social behavior. We first perform a comprehensive analysis of behavior in three economic games, finding evidence of discrimination against a sample from outside prison. In addition, our regression analysis reveals that inmates generally become less pro-social towards this out-group the longer they remain incarcerated. Finally, we introduce and evaluate a priming intervention that asks inmates to reflect on their time spent in prison. This intervention has a very sizeable and significant impact, increasing pro-sociality towards the out-group. Hence, a simple, low-cost intervention of this sort can have desirable effects in promoting rehabilitation and integration into social and economic life after release.

**JEL codes:** C93, K14, K42
**Keywords:** Prison inmates; social behavior; rehabilitation; experiment
**Declarations of interest:** none

# 1. Introduction

Fighting crime is among the priorities for any country's internal affairs. Crime is not only detrimental in terms of more insecurity and a lowering of a country's social capital, but it also represents a high direct monetary cost to society. For instance, in the United States, the annual cost of crime is estimated in 3 trillion US dollars (Anderson, 2012). The number of incarcerated people in the world is more than 10 million, with more than 2.2 million in the US. In terms of prison population rates, in the US and in Western European countries there are 698 and 84 prisoners per 100,000 of national population, respectively. Overall, the world prison population has increased by around 20% since 2000 (Walmsley, 2016). This increase is driven by Central (over 80%) and South America (145%) and to a lesser extent by the US (14%), while a decreasing pattern (by 21%) is seen in Europe. It is also important to highlight the huge recidivism rates, which may explain part of the overall increasing rates for crime. US data show a non-encouraging recidivist pattern: around two-thirds of ex-convicts were rearrested within three years, and three-quarters within five years. Among those who were rearrested within five years, 36.8% were arrested within the first six months (Durose et al., 2014).

In light of the above patterns, it seems crucial to gain a better understanding of how prison sentences affect inmates, and to examine ways to promote their rehabilitation and re-integration into social and economic life. As a matter of fact, rehabilitation and social reintegration are – along with retributive justice and deterrence – among the main objectives of modern penal systems. According to the European Court of Human Rights, '*while punishment remains one of the aims of imprisonment, the emphasis in European penal policy is now on the rehabilitative aim of imprisonment'*.[1] But how successful are prison sentences in fostering pro-social behavior and promoting the future integration of inmates into social and economic life? The debate among criminologists about whether prisons have a positive effect on deterring inmates' criminal behavior remains unresolved. Generally speaking, different schools of thought have expressed contradictory points of view in this regard (Gendreau et al., 1999). The first one is optimistic and claims that, in line with the concepts of specific and general deterrence, prison sentences suppress criminal behavior. In support of this view and in particular specific deterrence, recent evidence from Norway points towards a preventive effect of time spent in prison on subsequent criminal activity and recidivism by individuals and within their family and criminal network (Bhuller et al. 2016, 2018). The second goes in the exact

[1] See *Vinter* [GC], supra n 1 at para 115. Here the Grand Chamber referred to *Dickson v United Kingdom*, supra n 35 at para 75; and *Boulois v Luxembourg* ECHR Reports 2012; 55 EHRR 32 at para 83.

opposite direction, claiming that prisons promote criminality given the psychologically destructive and inhumane conditions (Chen and Shapiro, 2007) and the possibility of acquiring new criminal skills from one's fellow inmates (Bayer et al., 2009). Hence, the discussion over the effect of prisons on inmates remains open.[2] In any case, more work is needed to understand inmates' social behavior in order to design effective policies to prevent or rehabilitate criminals.

While this paper is not equipped to provide a conclusive answer on how prison sentences affect behavior, it contributes to this discussion by measuring important aspects of inmates' social behavior and by evaluating a simple intervention aimed at increasing pro-sociality. We present evidence from a series of experiments based on questionnaires and simple games that have been extensively used in the economic literature in order to extract various aspects of social behavior, namely the Trust Game, the Prisoner's Dilemma, and the Equality Equivalence Test that elicits social preference types (henceforth TG, PD, and EET, respectively). The experimental design allows us to address the following topics, which can be informative to the relationship between imprisonment and social behavior: First, we perform a comprehensive analysis of inmates' social behavior along the dimensions of pro- and anti-social distributional preferences, trust and reciprocity, and cooperation. Second, we search for parochial altruism and intergroup discrimination by comparing the behavior of prison inmates towards the in-group (other inmates) and the out-group (a sample from outside the prison). We believe this is particularly important, given that successful rehabilitation and re-integration will require inmates to interact with the out-of-prison population in the future: in that sense, discriminatory behavior against the (current) out-group can be worrisome and indicative of unfavorable attitudes towards the general population. Finally, our study introduces and evaluates a simple psychological priming method at the beginning of the experimental sessions with the aim of activating a positive rehabilitation effect: we ask inmates to reflect on their time spent in prison and on how it has affected them. The idea behind this intervention is that inmates may be driven to think about the rehabilitative purpose of imprisonment and maybe express

[2] Some related empirical work has been carried out and provides evidence on the relationship between prison conditions and recidivism (Roodman, 2017; Villettaz et al., 2014). For instance, it has been shown in Drago et al. (2011) that prison severity does not reduce the probability of recidivism. Instead, it seems that harsh prison conditions increase post-release criminal activity. Further work finds little correlation between the length of sentences and employment or income (Toch, 1977; Needels, 1996). Experiments have been also conducted in prison to study the criminogenic and anti-social behavior of prisoners as well as the effectiveness of rehabilitation programs (e.g., Farrington and Welsh 2005; Lipsey and Cullen, 2007). Finally, a few controlled experiments evaluate the effect of prison sentences on outcomes such as employment or re-offense rates, by comparing short-term imprisonment against community service (Killias et al, 2000, 2010).

through their decisions in the experimental games a desire or a commitment to become 'better persons'.

Our methodology consists of a lab-in-the-field experiment (Harrison and List, 2004) with a non-standard subject pool of 105 prison inmates from two prisons, a low and a high-security one, in Chania, Greece. We complement our experimental data with administrative and survey data on prison inmates. Our choice of games (TG, PD, EET) is motivated by the fact that the elicited facets of pro-sociality are essential in the time after release for prison and keys to a successful rehabilitation. Trust and reciprocity are of particular importance, not only in social relationships, but also for professional relationships that are commonly based on mutually beneficial gift-exchange (Akerlof, 1982; Fehr et al., 1997). Similarly, many jobs require workers to cooperate with each other in teams, and cooperation has been put forward as the most essential and defining feature for the organization of modern societies. Importantly, economic experiments have demonstrated that cooperation rates measured in standard economic games correlate with actual behavior in the field (Fehr and Leibbrandt, 2011; Gneezy et al., 2016). Finally, distributional preferences elicited in the EET – such as altruism or inequality aversion – have obvious implications for successful social exchange, but can also affect domains such as workplace behavior (Putterman, 2006; Rotemberg, 2006) or voting decisions (Fisman et al., 2017).

The literature in experimental economics studying the behavior of inmates, outlined in the next section, is surprisingly limited. Personality inventories and interviews have been extensively used by psychologists and criminologists to assess certain aspects of prisoners' personality and individual attitudes (Toch, 1977). However, their social preferences and behavior in interaction within a social group have not been elicited in a systematic and controlled way in incentive compatible contexts where inmates could interact either with each other or with non-institutionalized subjects. Moreover, we are aware of almost no controlled economic experiments sharing our aims of assessing the effect of prison sentences on various aspects of social behavior (other than post-release criminal activity measured by recidivism rates), and proposing ways of fostering a positive rehabilitation effect such as our priming intervention. One exception is a recent paper examining preference stability by evaluating the effect of participation in an 'offender accountability program' aimed at promoting pro-social behavior in a sample of Californian inmates (Maggioni et al., 2018). The findings indicate that

participation in this ten-month long program increased trust (but not altruism), hence providing an example of a successful intervention in the direction of rehabilitation.[3]

Our results reveal that inmates exhibit some degree of discrimination in their behavior: they are significantly less reciprocal and less cooperative towards the out-group (people from outside the prison) than towards the in-group (other inmates). However, discrimination is not present in the sample of inmates that experienced our priming intervention, because the intervention has a very large and statistically significantly positive effect on pro-sociality towards the out-group in the domains of trust, reciprocity, and cooperation. These findings suggest that a simple, low-cost intervention based on encouraging prison inmates to reflect on their incarceration may be an effective tool to promote rehabilitation and a more successful integration into social and economic life after release from prison. This intervention could easily be integrated into, e.g., counseling sessions routinely offered to prison inmates by trained psychologists or social workers. Furthermore, the regression analysis allows us gain some insights on how the time spent in prison affects pro-sociality: we find that longer time spent in prison is associated with lower cooperation rates towards the out-group, lower benevolence when lying ahead of another participant regardless of group, and stronger reciprocity towards the in-group. These findings can be useful when assessing the rehabilitation effect of prison sentences.

## 2. Literature review

### 2.1. Economic experiments in prisons

Prison inmates are a population that has only recently attracted the interest of experimental economic research, and only to a very limited extent. Running economic experiments in prisons can be challenging, as it requires complicated logistics. However, we believe that economists and behavioral scientists have a word to say on the topics we focus on here, and that insights from controlled experiments can be of major importance for criminologists and policy makers. In this section we present some related literature on the topics of social behavior among prisoners, priming, and ostracism.

[3] An interesting literature in criminology has used randomized trials in order to evaluate several kinds of interventions (such as boot camps, cognitive-behavioral interventions, and therapeutic programs for drug addicts and sex offenders) aimed at improving correctional outcomes and reducing recidivism rates. These interventions have generally been found to produce at best modest results. A summary is provided in Farrington and Welsh (2005).

Existing studies have mainly focused their attention on describing the social behavior of inmates. Using a dictator game, Birkeland et al. (2014) find no significant differences in altruism and no evidence of discrimination between criminals and non-criminals. Additionally, in an internet experiment, these authors do not find any difference in the pro-social behavior between people with a criminal record and people without. The authors claim that pro-social motivation captured by the dictator game is not a sufficient component to explain criminal propensity. Similarly, Chmura et al. (2010) find no significant difference in the propensity to give in a dictator game between prisoners and an average group of subjects. Those prisoners who attend prison school are less likely than the control group to share their endowment, but if they do share it, they give considerably more. Particularly those inmates who have the better marks and especially those who improve their marks over time are those who give more. To some extent contradicting the above findings, Chmura et al. (2016) find that prisoners give more than students in a dictator game and they give more to charity than to another prisoner. Our findings complement these studies since we use a new method to elicit distributional preferences, which does not rely on a one-dimensional index of pro-social motivation, but instead uses a choice list in order to classify decision makers into different types (for more details on this method, see section 3).

On the topic of cooperative behavior, Khadjavi and Lange (2013) compare female inmates to a sample of students in a simultaneous and a sequential prisoner's dilemma. Interestingly, in the simultaneous version inmates cooperate more than students do. Clark et al. (2015) run four cooperative games in the US. They compare their behavior to a sample of the general population. Overall, their results show that offenders make lower offers in the ultimatum game, cooperate less in the PD, display lower trust and reciprocity in the trust game and make lower contributions in the public good game.[4] Nese et al. (2018) run an experiment in order to provide evidence on cooperation and response to sanctions among different types of criminals. They conduct a PD and a third-party punishment game with three different samples: students, 'ordinary criminals' and Camorristi (Neapolitan 'Mafiosi'). Their results show that Camorra prisoners have a high degree of cooperativeness and a strong tendency to punish free-riders, as well as a clear rejection of the imposition of external rules even at significant cost to themselves. Ordinary criminals are the least cooperative group of participants in this study,

[4] It must be noted, however, that the procedures do not meet the standards of experimental economics since subjects did not know the identity of the other player who in the end turned out to be the experimenter.

lagging behind Camorra prisoners as well as students. On the other hand, ordinary criminals and Camorristi punish more than students in the third-party punishment game.

Khadjavi (2015) tests the effectiveness of deterrence in a German prison by conducting two treatments of the so-called *Stealing Game*. In the first treatment, subjects have the option to steal without consequences, whereas in the second treatment they face a potential punishment when stealing. His main results show that inmates are less willing to steal when they face the possibility of punishment, suggesting that deterrence can mitigate stealing behavior. Overall, the studies discussed above suggest that inmates exhibit a largely similar social behavior to that of students or groups drawn from the general population in non-strategic settings such as the purely distributional choice in the dictator game, while behavior in more interactive games differs to varying degrees.

An interesting recent study by Cohn et al. (2015) uses the coin tossing task (Bucciol and Piovesan, 2011) in a Swiss prison to study the effect of priming on inmates' ethical behavior. The results show that inmates cheat more when they are primed by making their criminal identity salient. The authors contrast this result with a control group of non-criminals where they do not find an effect of priming. Hence, criminals are found to behave differently when their criminal identity is activated. Moreover, Cohn et al. (2015) show that cheating rates in the experiment correlate with the number of disciplinary offences in prison, lending strong support to the external relevance of the experimental findings and the methodology used.

### 2.2. Priming

Economists have in recent years increasingly adopted priming techniques that are commonly used in psychology experiments.[5] Priming consists of subtly activating certain mental reminders to manipulate subsequent decisions by experimental participants (Bargh and Chartrand, 2000). The most common priming methods used by economists include questionnaire content manipulation, word unscrambling tasks, paragraph reading, background images, norm salience using a confederate, counting tasks, visual or audio cues, or writing a paragraph on a given topic. To give a few examples, Cohn et al. (2015) primed financial professionals with either a boom or a bust scenario. In treatment 'Boom' subjects saw an animated, fictive chart of a booming stock market while in the bust condition they saw a chart corresponding to the bursting of a bubble. Cohn et al. (2014) primed bankers by using questionnaires that made their professional identity salient; the questionnaire method was also

[5] For a review on priming in economics, see also Cohn and Maréchal (2016).

used in the paper by Cohn et al. (2015) run in Swiss prisons and discussed earlier. Shariff and Norenzayan (2007) as well as Benjamin et al. (2016) used a word unscrambling task in order to induce a religious prime. Gino et al. (2009) primed participants by making cheating appear as a salient social norm. Some studies have found that using subtle cues that make subjects feel observed increases pro-sociality (Bateson et al., 2006; Haley and Fessler, 2005; Ekström, 2012), while others (such as Fehr and Schneider, 2010) have failed to find an effect. Andersson et al. (2017) used subliminal messages of a pro-social prime and found an increase in donations to charity. Vohs et al. (2006) primed subjects with reminders of money, which led them to request less help and to be less helpful towards others. Callen et al. (2014) conducted an experiment on risk decision in Afghanistan, priming subjects by asking them to recall and recount fearful experiences.

### 2.3. Ostracism

We finally mention a few economic experiments that have studied the effect of ostracism in social dilemmas. This literature is related to our work because imprisonment can be seen as a form of ostracism: criminals are excluded from social life and deprived of their rights and social benefits for a certain period of time. This strand of the literature has mainly focused on the effect of the threat of exclusion on cooperation and contribution to public good games (Masclet, 2003; Cinyabuguma et al., 2005; Kerr et al., 2009; Sheremeta et al., 2011; Maier-Rigaud et al., 2010). Overall, the results provide evidence that the threat of exclusion reduces free-riding behavior. Other studies have focused on the type of exclusion, i.e., whether it is irreversible (Cinyabuguma, et al., 2005; Maier-Rigaud et al., 2010) or whether subjects are free to move immediately to another group (Charness and Yang, 2014).

More closely related to our work, the lab experiment of Solda and Villeval (2017) analyses the behavior of excluded members, after their reintegration, into a public goods game. Subjects are excluded for a determined period of time and not irreversibly. The authors manipulate the possible length of exclusion and whether this is determined exogenously or as a result of a vote. As in the aforementioned studies, subjects typically exclude the less cooperative group members. After reintegration, excluded members increase cooperation rates and compliance with the group norm but only when the exclusion is imposed for a short period of time, endogenously chosen by vote. Hence, a short duration of exclusion can be advantageous in promoting reintegration. This result is particularly interesting for us, since it concurs with our finding that a longer time spent in prison is associated with lower pro-sociality towards the out-group.

# 3. Experimental design and hypotheses

Our experimental design is based on three simple games, supplemented by collection of sociodemographic data, questions related to inmates' experience inside the prison and data provided by the prison administration.

## 3.1. The games

**TG:** We use a discrete version of the trust game (Berg et al., 1995). Subjects are matched in groups of two and have the role of either player 1 (the sender) or player 2 (the receiver). The game is sequential with player 1 as first mover and player 2 as second mover. The sender has two strategies, to trust or not to trust player 2. If he decides not to trust, the game ends and both players earn 10 euros. If the sender decides to trust, then the game continues and player 2 has to decide whether to reciprocate or not. If he decides to reciprocate they both earn 20 euros, while if he decides not to reciprocate player 1 earns 5 euros and player 2 earns 35 euros. Given this design we are able to classify subjects in the role of player 1 as trusting or not trusting, and subjects in the role of player 2 as reciprocal (trustworthy) or not reciprocal (not trustworthy).[6] For self-regarding, money maximizing subjects the subgame perfect equilibrium prediction in this game is straightforward. Reciprocating reduces the receiver's payoff; hence, the receiver will never reciprocate. Anticipating this, senders will not trust initially. On the contrary, the Pareto efficient solution that provides the highest payoffs is that player 1 trusts and player 2 reciprocates.

In our experiment, subjects were randomly assigned to the sender's or the receiver's role. We implemented the game in a simultaneous rather than in a sequential form, as follows: senders had to indicate whether they trusted the receiver or not, while at the same time receivers had to indicate whether they wanted to reciprocate in case the sender had decided to trust them. This strategy method ensured that we collected data on the reciprocity of all subjects who were assigned to the role of the receiver.

**PD**: We use the same version of the simultaneous prisoner's dilemma as Khadjavi and Lange (2013): Two players simultaneously decide either to cooperate with the other player or to defect. Assuming selfish own money maximization, the dominant strategy for both players

[6] It should be noted that splitting the surplus may be due to several reasons. It can indicate reciprocal behavior or a desire to reward trust, but it can also be driven by altruism or inequality aversion on part of the second mover. For thorough discussions on other-regarding motivations in the trust game see Cox (2004), Isoni and Sugden (2018). We are not able to identify the motivations driving the first- and second mover's behavior in the trust game, not least because we did not elicit participants' beliefs in the experiment.

is to defect and, therefore, the unique Nash equilibrium predicts mutual defection, yielding payoffs of 3 euros to both. If both players cooperate the payoffs are 7 euros to each player, which Pareto dominates the Nash equilibrium. If the players choose different strategies, the defector obtains 9 euros while the cooperator obtains 1 euro.

**EET**: The Equality Equivalence Test (Kerschbamer, 2015) elicits distributional preference types. Each subject has to make ten binary choices between allocations that both involve an own payoff for the decision maker and a payoff for a randomly matched anonymous second subject. In each of the ten binary decision problems, one of the two allocations is symmetric – i.e., egalitarian, giving the same payoff to each person – while the other leads to unequal payoffs for the two subjects. The ten choices are shown in Table 1, which breaks them down into a so-called disadvantageous inequality block and an advantageous inequality block, depending on whether inequality is to the advantage or disadvantage of the decision maker. In both blocks the symmetric payoff and the payoff of the second subject are held constant, while the payoff of the decision maker increases in steps of 40 cents. Given this design, in each of the two blocks a rational decision maker switches at most once from the symmetric to the asymmetric allocation (and never in the other direction).

Based on the choices in this task each subject reveals information about his or her benevolence in the domains of disadvantageous and advantageous inequality. We use a simple coding procedure that assigns values from -2.5 to +2.5 (in steps of 0.5) to revealed benevolence in each domain, with higher values corresponding to higher benevolence. Benevolence in the domain of disadvantageous inequality is thus measured by the so-called 'x-score' and benevolence in the domain of advantageous inequality by the so-called 'y-score'. Based on these values, we can also classify each subject into one of four distinct behavioral types. In particular, a decision maker who is benevolent in both domains is classified as an altruist; a decision maker who is benevolent when ahead but malevolent when behind is classified as inequality averse (IAV); a decision maker who is malevolent in both domains is classified as spiteful, and a decision maker who is benevolent when lying behind but malevolent when lying ahead is classified as inequality loving.[7]

[7] For more details on the precise way in which benevolence and malevolence towards the second subject are defined and measured we refer to Kerschbamer (2015). Note that, according to the classification used here, selfish subjects are a subset of the four other categories. An alternative classification method would include selfish subjects in addition to those classified as strictly altruistic, inequality averse, or inequality loving. Using this alternative classification does not change any of our main findings; details are available upon request.

**Table 1: The Equality Equivalence Test (EET)**

**Benevolence Behind**

| **LEFT** | | | | **RIGHT** | |
|---|---|---|---|---|---|
| You get | Another person | | | You get | Another person |
| 3.2 | 5.2 | **LEFT** | **RIGHT** | 4 | 4 |
| 3.6 | 5.2 | **LEFT** | **RIGHT** | 4 | 4 |
| 4 | 5.2 | **LEFT** | **RIGHT** | 4 | 4 |
| 4.4 | 5.2 | **LEFT** | **RIGHT** | 4 | 4 |
| 4.8 | 5.2 | **LEFT** | **RIGHT** | 4 | 4 |

**Benevolence Ahead**

| **LEFT** | | | | **RIGHT** | |
|---|---|---|---|---|---|
| You get | Another person | | | You get | Another person |
| 3.2 | 2.8 | **LEFT** | **RIGHT** | 4 | 4 |
| 3.6 | 2.8 | **LEFT** | **RIGHT** | 4 | 4 |
| 4 | 2.8 | **LEFT** | **RIGHT** | 4 | 4 |
| 4.4 | 2.8 | **LEFT** | **RIGHT** | 4 | 4 |
| 4.8 | 2.8 | **LEFT** | **RIGHT** | 4 | 4 |

### 3.2. Identity and in-group out-group effects

The term *identity* is used here as an individual's self-perception regarding his or her main characteristic(s) which define(s) them as a person. In general, identity may be driven by a person's individual features (such as a family history[8]) and exceptional endowments or achievements (the tallest man in the world, the best student in class, feeling of supernatural or paranormal sensations, etc.), or by a feeling of belonging to a particular group (national, religious, linguistic, racial, geographical, consumption of a common brand, etc.). [9]

One manifestation of group identity may be in-group or parochial favoritism, understood as the tendency to favor members of one's own group over those in other groups (Everett et al., 2015). In fact, the discrimination comes from two independent but possibly co-evolving tendencies, namely, cooperation within and competition between groups (Rusch, 2014). Based on this reasoning, we derive our first testable hypothesis:

[8] See, for example, Phelan and Rustichini (2018).

[9] Literally, "*that part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership*" (Tajfel, 1974, p. 69).

*H1: Inmates discriminate against the out-of-prison sample.*

This hypothesis relies upon the assumption that a prisoner's identity is salient among the prison sample. As we will show in section 4.2, this is indeed the case. However, we must mention that differences between the inmates' in- and out-group behavior could also be due to things that are unrelated to group membership, such as differences in perceived income or deservingness.

In order to test H1, we let participants in our experiment make their decisions in each game under two conditions in a within-subjects design. In the in-group condition, prison inmates know that they are matched with another inmate, while in the out-group condition inmates know that the partner is someone from outside the prison. Non-institutionalized subjects know that in the in-group condition the partner is someone from the same school, while in the out-group condition the partner is someone who is in prison. We control for order effects by randomizing the order in which participants make their decisions for the in-group and the out-group. Asking participants to decide twice and varying the partner's identity has been used by several studies on discrimination (see, e.g. Birkeland et al., 2014; Chen and Li, 2009; Romano et al., 2017), and an advantage of this design choice is that it allows us to study our research question with the full sample size and without compromising statistical power. However, this methodology also faces limitations, and it can be criticized for being prone to experimenter demand effects (Zizzo, 2010).[10]

### 3.3. Priming

In a between-subjects design, inmates are randomly assigned to either the neutral (control) or the priming treatment. In the priming treatment, participants are asked to write a piece of text on the following topic: "*Reflect on your time spent in prison, tell us how it has affected your social behavior*". In the neutral treatment, inmates write a paragraph based on the following: "*Reflect on an activity you like, tell us how it has affected your life*". The procedures are very simple: at the beginning of each session and before making their decisions in the subsequent

[10] In this respect it is worthwhile noting that administration in both prisons has confirmed to us that the inmates have had no previous experience with research studies inside the facility, thus reducing the risk that they are familiar with experimental procedures and hypotheses.

economic games (TG, PD, EET), subjects receive a sheet with the corresponding text and a blank space to write the paragraph.[11]

In such a context, group reputation concerns are likely to emerge and affect individual behavior, in a similar way in which social dilemmas emerge from the conflict between the common and the private interest. Everett et al. (2015) argue that a person who identifies with a group may act in a way resembling a social dilemma, where the public good may be the group's reputation. Interestingly, it has been shown (Hopkins et al., 2007) that, in the name of his or her group's reputation, a subject may choose more cooperative actions towards the out-group in order to induce out-group members to see in-group members in a more positive way. This argument motivates our second testable hypothesis:

*H2: Inmates in the priming treatment display stronger pro-sociality towards the out-of-prison sample, compared with those in the neutral treatment.*

This hypothesis rests on the assumption that priming increases the salience of the prisoner's identity. In section 4.2 we will present evidence in support of this assumption.

We have conducted a post-experimental content analysis of the inmates' accounts in the priming treatment, in order to form an impression of the nature of these accounts. In particular, we asked two independent research assistants who were blind to our hypotheses to code each account as conveying a generally positive, generally negative, or neutral message. The codings of the two assistants are very highly correlated (Kendall's $\tau = 0.66$, $p < 0.01$; tau-equivalent reliability = 0.80) and reveal strong differences between the neutral and primed group. [12] While the large majority of accounts (35 of 47, or 74.5%) in the neutral group were coded as neutral in terms of content, neutral accounts represent only 25.9% of the total (14 of 54) in the primed group. This difference is statistically significant ($\chi^2 = 23.71, p < 0.01, N = 101$) and supports the validity of our priming tool, by showing that we have indeed succeeded in creating a neutral group against which the effect of the priming instrument can be evaluated.

To summarize, subjects have to decide sequentially in six games: TG in-group, TG out-group, PD in-group, PD out-group, EET in-group, and EET out-group. Therefore, we conduct three economic games in a 2 x 2 factorial design combining the in-group vs. out-group

[11] We implement the priming manipulation only with the sample of prison inmates, given that the priming condition requires participants to reflect on and give an account of their time spent in prison: this would not be a reasonable task for students who have never been imprisoned.

[12] The content analysis has been conducted for 101 out of 105 inmates' accounts, since four accounts were illegible. In all cases of disagreement (N=24), the two coders had to meet and decide on a common coding.

manipulation (within-subjects) and the neutral vs. priming manipulation (between-subjects). Note that we randomly assigned participants as either player 1 (trustor) or player 2 (trustee) in the TG, hence we have 59 inmates in the role of player 1 and 46 as player 2.

### 3.4. Questionnaires

After making their decisions in the TG, PD and EET, inmates were asked to fill out a questionnaire about their nationality, age, marital status, education level, and number of siblings (see Appendix B). Additionally, we asked them to answer questions regarding the conditions of their imprisonment: time spent in the current prison, number of times imprisoned, total time spent in prison during their life, type and length of sentence, attendance of religious activities in prison, number of cell mates, frequency of leaving the prison (for any reason) and number of working days per month. The prison administration provided us with this same information, allowing us to double check and correct for minor discrepancies.

### 3.5. Procedures with institutionalized subjects

All sessions were run in two male prisons in Chania, Greece. In November 2016 we ran one session in the high security prison facility "Crete 1" and two simultaneous sessions in the low security agricultural prison facility of "Agia". In April 2017 we conducted an additional session in the low security prison. We recruited a total of 105 volunteer male inmates by posting announcements in several languages (Greek, English, Arabic and French), around the prison premises to invite inmates to participate in the experiment. Additionally, two days before each session, the experimenters went to the prison to answer possible questions and give a short explanation of what is an economic experiment. Once they decided to participate, inmates had to register through the prison administration.

In the high security prison, the single session was conducted in the library (N=14). In the low security prison, the two simultaneous sessions were run in the library and in the gym (N=31 and N=12, respectively), and the last session was run only in the library (N=48). No guards were present during the sessions and we insisted on and guaranteed subjects about anonymity, by giving them a random number so there was no way to associate a decision with a name. [13] We were very cautious in minimizing any kind of audience effects.

[13] The administration was in charge of creating a double blind process between the experimentalists and the inmates. The reason to have this blind process was not only to ensure anonymity in the experimental procedure, but also to comply with the regulations of the penitentiary center. Each participant was given a random number, which was written in his booklet set. Once the inmate finished filling out the booklet and before leaving the

The experiment was conducted with pencil and paper. Subjects could choose among four different languages for their booklet of instructions: Greek, English, Arabic or French.[14] We enforced the usual experimental practice of not allowing for communication among subjects and ensuring anonymity in decision making. For that purpose, we had subjects seat in big tables of four with sufficient distance from each other, and we tried to seat them in mixed groups of different nationalities and languages of instructions to make communication and copying even less likely.

Once the session was ready to start, one of the experimenters explained aloud the general instructions of the experiment and answered possible questions. We stressed that the experiment was incentive compatible, meaning that participants would be paid with real money based on their decisions. Subjects were told that one game would be chosen randomly by the social worker at the end of the session. All the participants were invited to assist to the random draw of the game as witnesses. Given that inmates are not allowed to receive money directly, we explained to them that their payment would be credited to their personal prison account, which can be used to buy goods inside the prison (e.g., in the cafeteria).

Afterwards, the experiment started and participants were asked to keep silent until the end of the session. We randomized the order in which the PD and TG were presented and played, although we kept the EET always as the third game. The instructions for each game were read in silent by each subject and they could go through the booklet at their own pace. Before proceeding to the three games, all participants had to complete the priming task by writing the corresponding piece of text. Three experimenters were present in each session in order to answer any question in private and to assist participants. Whenever participants had finished filling out their booklet of instructions (including the priming text, the three games, and the questionnaires), they were allowed to leave the session.[15] We executed the payments

laboratory, he announced his random number to the administration, who recorded the number and associated it to the corresponding person. Afterwards, the administration provided us with additional data associated to each number. In this way we were only able to match a random number to the administrative data and experimental choices. Moreover, we did not provide information on inmates' experimental choices to the administration, to ensure anonymity between inmates and administration in this dimension and to respect inmates' privacy. This procedure was clearly explained to inmates, to ensure that their decisions were made free of any concerns about lack of privacy or anonymity.

[14] Instructions in languages other than English were translated from English by native speakers. In Table A1 (Appendix A) we show how many inmates chose to receive the instructions in each language. That table also shows the distribution of languages in which inmates completed the priming task. Six subjects chose to complete the task in a language other than Greek, English, Arabic and French. We note that our research team made sure that every participant understood the instructions, and to the best of our knowledge there were no cases where language posed a threat on understanding. Sample instructions can be found in Appendix B.

[15] We note that all participants completed all parts of the experiment and none left a session before doing so.

one day later, after randomly matching the decisions of the participants and calculating the resulting payoffs.

### 3.6. Procedures with non-institutionalized subjects

The non-institutionalized subjects were recruited in the Mediterranean Agronomic Institute of Chania (MAICH) also located in Chania, Greece. This is a boarding school in which students spend most of their time in a closed environment, hence they attend lectures, eat and sleep in the school, and typically share rooms with other students. We ran one session with 40 volunteer students from different degrees in November 2016. The procedures were as similar as possible to the ones followed in prison. Each subject had a booklet (in English, since this is the language of instruction at the school) with a different order between PD and TG and with different orders for the in-group and the out-group decision. Each subject was assigned a random number to ensure anonymity. One of the experimenters explained aloud the general instructions and answered possible questions. Afterwards, all subjects went alone through the different games and questionnaires. Similarly, subjects were aware that only one game would be randomly chosen at the end of the session to determine their earnings.[16] It is important to acknowledge that the sample of non-institutionalized participants differs to prison inmates along several dimensions, such as gender and age composition, educational background and socioeconomic background. The main purpose of collecting data from this sample is thus not to make direct comparisons between the behavior of the two samples, but rather to create an out-group that is used for matching purposes in the experiment. Nevertheless, for completeness, in Appendix A we show how the behavior of our focus group of inmates compares to that of students.

## 4. Results

We present our results in three subsections. We first describe our sample. Second, we show a landscape of subjects' social behavior and study in-group-out-group bias among inmates by documenting the differences in their social behavior when they are matched with another inmate versus with someone from outside prison. Third, we demonstrate the effect of the priming intervention on inmates' social behavior.

[16] Given that the number of subjects was higher in the prison than in the student sample, we matched each student to two or three prisoners in order to calculate payoffs. However, for each student we only paid out one of the matchings.

### 4.1. The sample

In the following analysis we pool the data from the low and the high security prisons, given that no significant differences in decisions were found between the two samples in any of the three experimental games. Our sample is composed of 145 subjects in total, 105 of whom are male inmates and 40 (25 female and 15 male) are subjects from the student group. Among the 105 inmates, 57 participated in the priming treatment and 48 in the neutral treatment. 59 inmates were randomly assigned to the role of player 1 and 46 the role of player 2 in the trust game.

Table 2 shows some descriptive statistics for the sample of prison inmates. In the upper panel of the table we provide information on the type of crime committed, classifying crimes into eight categories as recorded – and provided to us – by the prison administration: Drug related offences, robbery, smuggling goods, human trafficking, homicide, sexual crime, economic crime, firearms violations. Note that some inmates have been sentenced for two or more crimes. The variable Penalty (also provided by prison authorities) captures the total sentence in years. We also collected some information regarding inmates' living conditions in prison, including: Frequency of getting out of the facility (measured in days per month), number of people they share a cell with, number of days per month that they work inside the facility, and average number of visits per month. Finally, the lower panel of the table provides information about sociodemographic variables including mean age, marital status, number of children, number of siblings, and the level of education (coded as 0: none; 1: elementary 2: secondary school; 3: high school; 4: university; 5: master). The table breaks down all variables between the neutral and priming group, and shows that our sample is quite balanced among the two treatments, albeit with a number of exceptions. In the regression analysis of section 5 we therefore control for a number of potentially relevant variables.

**Table 2: Descriptive statistics for the sample of inmates**

| | FULL SAMPLE | | | NEUTRAL TREATMENT | | | PRIMING TREATMENT | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Rel. freq. | Std. dev. | N | Rel. freq. | Std. dev. | N | Rel. freq. | Std. dev. | p-value |
| **Type of** | | | | | | | | | | |
| Drug related | 43 | 0.41 | 0.49 | 19 | 0.40 | 0.49 | 24 | 0.42 | 0.50 | 0.79 |
| Robbery | 36 | 0.34 | 0.48 | 18 | 0.37 | 0.49 | 18 | 0.32 | 0.47 | 0.52 |
| Smuggling | 6 | 0.06 | 0.23 | 3 | 0.06 | 0.24 | 3 | 0.05 | 0.22 | 0.83 |
| Human | 13 | 0.12 | 0.33 | 5 | 0.10 | 0.31 | 8 | 0.14 | 0.35 | 0.57 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Homicide | 7 | 0.06 | 0.25 | 2 | 0.04 | 0.20 | 5 | 0.09 | 0.28 | 0.35 |
| Sexual crime | 3 | 0.03 | 0.17 | 1 | 0.02 | 0.14 | 2 | 0.03 | 0.19 | 0.66 |
| Economic | 8 | 0.08 | 0.27 | 3 | 0.06 | 0.24 | 5 | 0.09 | 0.28 | 0.63 |
| Firearms | 7 | 0.06 | 0.25 | 4 | 0.08 | 0.28 | 3 | 0.05 | 0.22 | 0.53 |

| Variable | N | Mean | Std. dev. | N | Mean | Std. dev. | N | Mean | Std. dev. | p-value |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sentence and prison conditions** | | | | | | | | | | |
| Penalty | 101 | 14.81 | 18.2 | 45 | 11.20 | 14.17 | 56 | 18.18 | 21.5 | 0.09 |
| Time served | 105 | 3.87 | 4.79 | 48 | 3.55 | 4.58 | 57 | 4.15 | 4.98 | 0.29 |
| Frequency out | 105 | 4.07 | 32.3 | 48 | 7.60 | 47.63 | 57 | 1.08 | 4.96 | 0.72 |
| Cell shared | 103 | 1.60 | 1.59 | 47 | 1.74 | 1.49 | 56 | 1.43 | 1.52 | 0.29 |
| Work days | 105 | 95.89 | 171. | 48 | 95.46 | 145.6 | 57 | 96.26 | 191. | 0.69 |
| Visits | 91 | 0.27 | 0.45 | 44 | 0.27 | 0.45 | 47 | 0.28 | 0.45 | 0.97 |
| **Sociodemographic variables** | | | | | | | | | | |
| Age | 105 | 35.95 | 9.40 | 48 | 33.65 | 8.60 | 57 | 38.00 | 9.68 | 0.02 |
| Married | 104 | 0.37 | 0.48 | 48 | 0.25 | 0.44 | 56 | 0.46 | 0.50 | 0.02 |
| Siblings | 104 | 4.83 | 4.15 | 47 | 4.57 | 2.88 | 57 | 5.04 | 4.98 | 0.58 |
| Education | 101 | 2.00 | 0.92 | 48 | 2.00 | 0.95 | 53 | 2.00 | 0.90 | 0.88 |

The upper panel reports relative frequencies of each type of crime in the subject pool. Note that frequencies do not add up to 1 because some of the participants have two or more sentences. The lower panel reports means of selected variables. The p-values correspond to t-tests for continuous variables, and chi-squared tests for binary variables and Education, comparing Neutral with Priming treatment.

### 4.2. In-group-out-group bias: Inmates' parochialism and discrimination

In this section we test hypothesis H1, by providing a description of inmates' behavior in the three experimental games and searching for parochialism and discrimination by comparing behavior towards the in-group (other inmates) and towards the out-group (people from outside prison). Figure 1 shows trust and reciprocity rates in the TG. The levels of trust are very similar: 44.07% of inmates trust the in-group versus 40.68% who trust the out-group (McNemar $\chi^2 = 0.02, \mathrm{p} = 0.89, \mathrm{N} = 118$). On the other hand, we observe a large difference in terms of reciprocity, i.e., trustworthiness by second movers: 65.22% of them are reciprocal towards the in-group and split the available surplus, versus 30.43% who share with the out-group. This difference is significant (McNemar $\chi^2 = 5.11, \mathrm{p} = 0.02, \mathrm{N} = 92$) and indicates discrimination against the out-group.

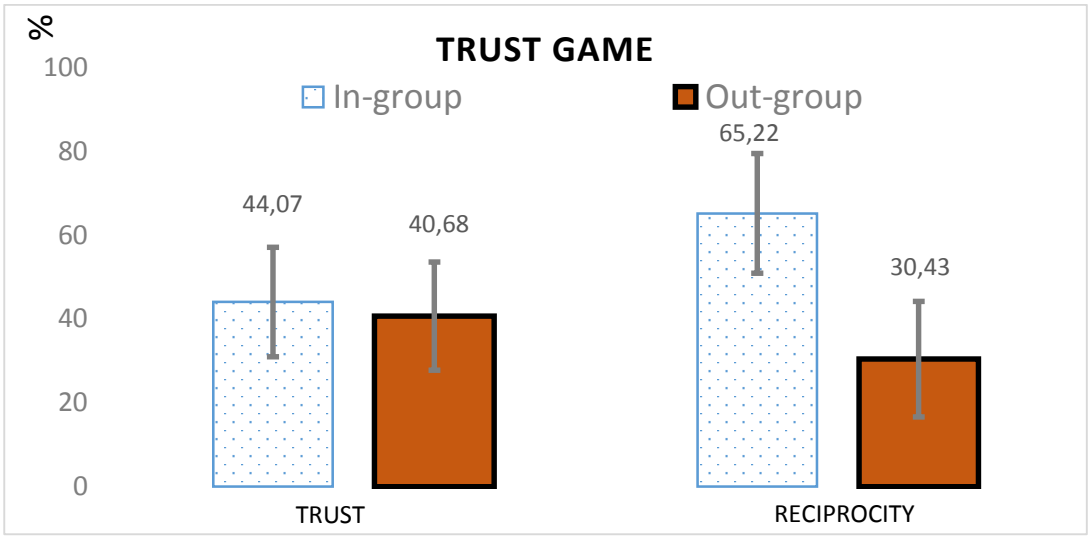


**Figure 1:** Inmates' decisions in the TG, broken down by behavior towards the in-group and out-group. Trust is measured as percentage of first movers who choose to trust in the TG; reciprocity is measured as percentage of second movers who choose to reciprocate in the TG. Bars include 95% confidence intervals.

Furthermore, we observe discrimination against the out-group in terms of cooperation. As shown in Figure 2, 55.24% of inmates cooperate with other inmates, and only 31.43% cooperate with people from outside prison. The difference is statistically significant (McNemar $\chi^2 = 6.33, \mathrm{p} = 0.01, \mathrm{N} = 210$).

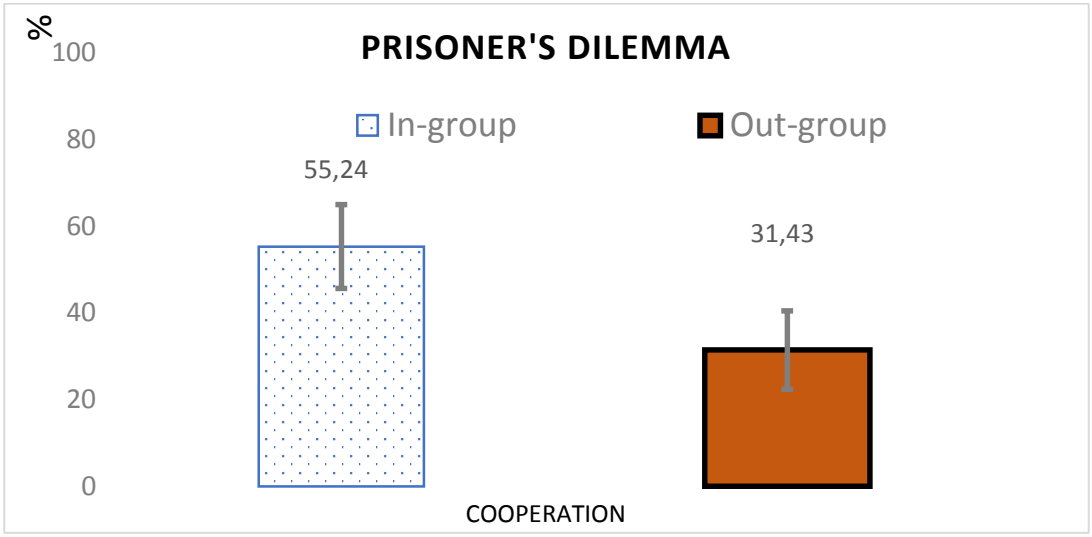


**Figure 2:** Inmates' decisions in the PD, broken down by behavior towards the in-group and out-group. Bars show cooperation rates and include 95% confidence intervals.

Figure 3 plots the distributional preference types among prisoners based on their decisions in the EET, again differentiating between decisions made to the in-group and to the

out-group. Overall, we do not find any significant differences in this regards. 20% and 19.05% of inmates are spiteful when interacting with another inmate and with someone from outside the prison, respectively (McNemar $\chi^2 = 0.001, \mathrm{p} = 1.00, \mathrm{N} = 210$). 13.33% and 15.25% are altruistic with the in-group and out-group, respectively (McNemar $\chi^2 = 0.03, \mathrm{p} = 0.85, \mathrm{N} = 210$). Around one third are inequality averse when interacting with both groups (McNemar $\chi^2 = 0.001, \mathrm{p} = 1.00, \mathrm{N} = 210$). Similarly, around one third are inequality loving when they are matched with another inmate and with someone from outside the prison as well (McNemar $\chi^2 = 0.01, \mathrm{p} = 0.90, \mathrm{N} = 210$).

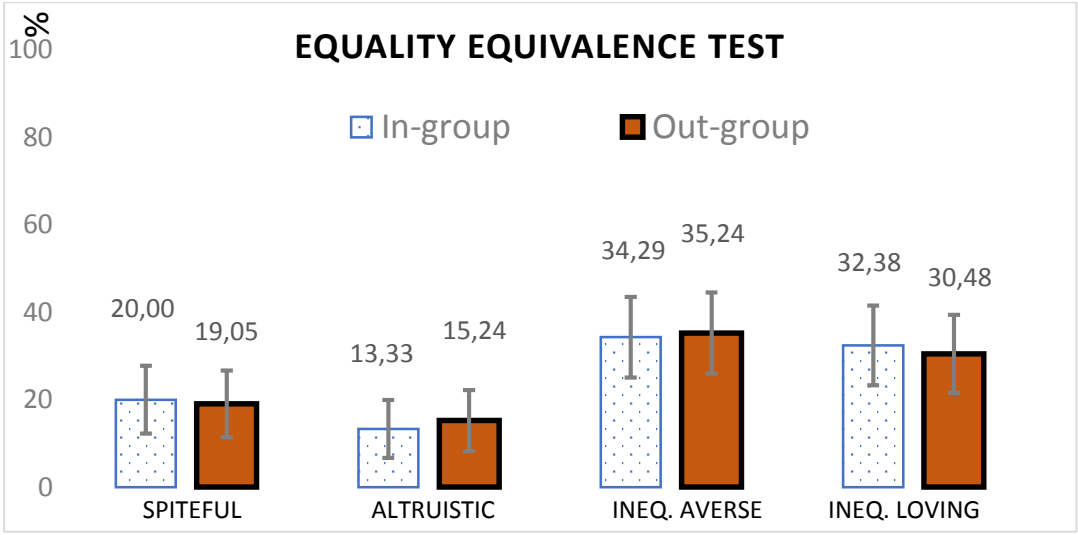


**Figure 3:** Inmates' classification of types (Spiteful, Altruistic, Inequality Averse, Inequality Loving) in the EET, broken down by behavior towards the in-group and out-group. Bars show the percentage of subjects classified into each type and include 95% confidence intervals.

The very high prevalence of inequality loving types, accounting for 31.43% of the total, is a striking feature of the distribution of types. This rate is more than six times higher than among students (see Appendix A for details), and it is particularly noteworthy since previous studies that use the EET report very low frequencies of this behavioral type (Balafoutas et al., 2012; Balafoutas et al., 2014; Kerschbamer et al., 2017). A possible explanation for this result might be related to the hierarchical systems that commonly exist in prisons, meaning that inmates are used to being either in the role of the kiss-up or in the role of the boss, thereby increasing benevolence towards subjects lying ahead and decreasing benevolence towards those who are behind. However, since we cannot present evidence other than casual observation during the sessions to substantiate this claim, we restrict ourselves to simply documenting this

interesting finding on the exceptionally high proportion of inequality loving types among prisoners.[17]

The aforementioned results partly support H1 and present novel evidence of discrimination from inmates in favor of their in-group and against their out-group, expressed by means of lower reciprocity and cooperation rates. On the other hand, we find no significant differences in trust rates or in the EET. This latter result is in line with Birkeland et al. (2014) who also find no evidence of in-group bias in the behavior of prisoners in a non-strategic context such as that of the dictator game (or the EET in our experiment).

*Result 1: Inmates exhibit significantly less reciprocity and are less cooperative towards the out-group (i.e., people from outside the prison) than towards other inmates.*

Hypothesis H1 rests on the assumption that prison inmates indeed perceive other inmates as an in-group, with which they feel identified. Moreover, H2 was motivated by means of stronger concerns for group reputation among primed inmates, which in turn suggests that group identification should be stronger in the primed group. In order to support these assumptions we present here evidence from a post-experimental questionnaire, which we conducted only in the second set of sessions in April 2017 (N=48). In this questionnaire we asked inmates to indicate the extent to which they feel identified with each of three possible groups, including the ethnic and religious group and the group of prisoners. As shown in Table 3, identification was strongest with the prisoner group, followed closely by the ethnic group. Identifications with both groups were significantly stronger than with the religious group. These results confirm that being a prisoner is indeed a salient identity and our subjects perceive other inmates as their in-group. The lower panel of the table compares identifications across priming conditions: in line with the reasoning behind H2, the prisoner identity increases in strength among the primed inmates who are asked to reflect on their time spent in prison, while the other two identities do not differ significantly by treatment.

[17] A further potential explanation could be related to poor comprehension of the EET by inmates. In this respect, we first note that we took enough time to answer questions in private, and ensure understanding of all experimental games during the sessions to the extent possible. Moreover, we use the inmates' level of education as a proxy for cognitive ability and perform a median split, comparing the distribution of types among inmates with above and below median educational attainment, failing to find significant differences between the two distributions. We also compute the correlation coefficients between education level and x- and y-scores and find that they are close to zero and insignificant ($\rho$=-0.08, p=0.16; $\rho$=0.07, p=0.22, for x-scores and y-scores, respectively).

**Table 3: Survey evidence on inmates' identification with various groups**

| | N | Mean | St. Dev. | Ethnical vs. Religious | Ethnical vs. Prisoners | Religious vs. Prisoners |
|---|---|---|---|---|---|---|
| Ethnical Group | 48 | 3.45 | 1.64 | | | |
| Religious Group | 48 | 2.63 | 1.39 | | | |
| Prisoner Group | 48 | 3.56 | 1.59 | z = 3.01<br>p < 0.01*** | z = -0.49<br>p = 0.62 | z = -2.60<br>p < 0.01*** |

| | Neutral Treatment | Priming Treatment | Ranksum M-W test |
|---|---|---|---|
| Ethnical Group | 3.62 | 3.29 | z = 0.82, p = 0.41 |
| Religious Group | 2.79 | 2.46 | z = 0.55, p = 0.58 |
| Prisoner Group | 2.96 | 4.17 | z = -2.59, p < 0.01*** |

Coding: 1= I **do not** feel identified with this group, 5= I **definitely** feel identified with this group

Regarding possible discriminatory behavior among the group of students, we find no evidence of discrimination in any of the dimensions in the experiment: students are equally trusting, reciprocal, cooperative, and benevolent towards their in-group of other students than towards their out-group of prison inmates, in the sense that we document no significant differences in any of these dimensions. Given that the focus of our work is on the behavior of inmates, we relegate to Appendix A the presentation of student behavior broken down by partner group (see Figures A4 to A6).

### 4.3. Effects of the priming intervention

We begin the analysis of how priming affects inmates' social behavior by comparing choices between the group of prisoners who were subjected to our priming intervention and those who participated in the neutral treatment. The analysis is shown in Figure 4 and reveals a strong increase on inmates' trust and reciprocity towards the out-group as a result of the priming intervention. Although priming has no significant effect on trust towards members of the in-group ($\chi^2 = 0.22, \mathrm{p} = 0.66, \mathrm{N} = 59$), it almost triples the percentage of inmates who trust members of the out-group ($59.4\%$ vs. $18.5\%$; $\chi^2 = 10.13, \mathrm{p} < 0.01, \mathrm{N} = 59$). Similarly,

priming does not affect the levels of reciprocity towards the in-group ($\chi^2 = 0.66, \mathrm{p} = 0.42, \mathrm{N} = 46$), but it increases the level of reciprocity towards the out-group by a factor of more than three ($44\% \; vs. \; 14.3\%$; $\chi^2 = 4.76, \mathrm{p} = 0.03, \mathrm{N} = 46$).

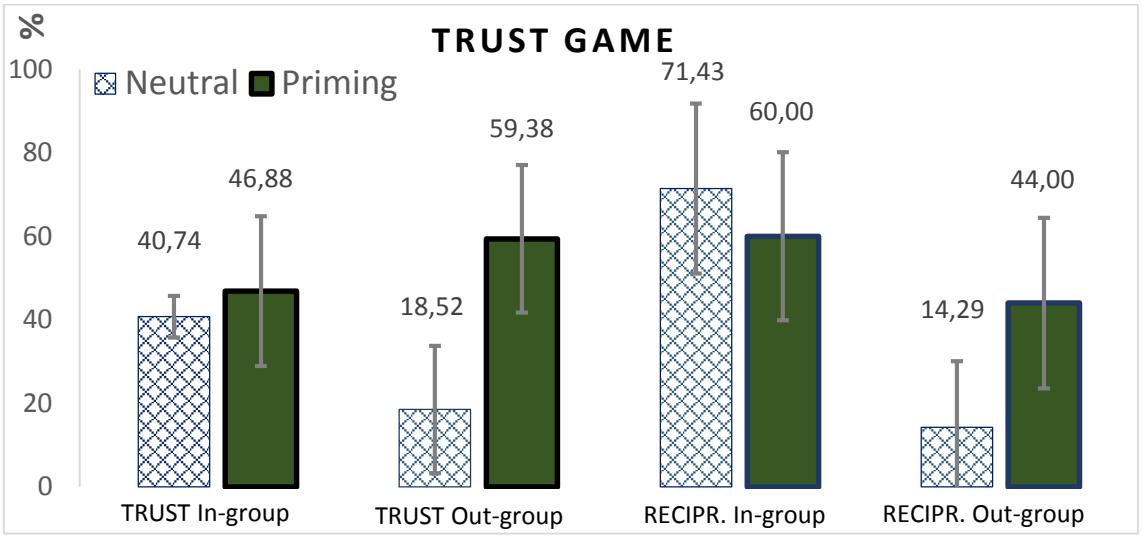


**Figure 4:** Inmates' decisions in the TG, neutral vs. priming treatment. 'Trust In-Group' and 'Trust Out-Group' denote trusting rates towards a partner from the in-group and out-group, respectively. 'Recipr. In-group' and Recipr. Out-group' define reciprocity rates in an analogous way. Bars include 95% confidence intervals.

This pattern is also present in the prisoner's dilemma. Figure 5 shows a non-significant effect of priming on cooperation towards the in-group ($\chi^2 = 1.89, \mathrm{p} = 0.17, \mathrm{N} = 105$), but a strong increase of cooperation towards the out-group ($38.6\% \; vs. \; 22.9\%$; $\chi^2 = 2.98, \mathrm{p} = 0.08, \mathrm{N} = 105$). Hence, hypothesis H2 is confirmed for the dimensions of trust, reciprocity, and cooperation.

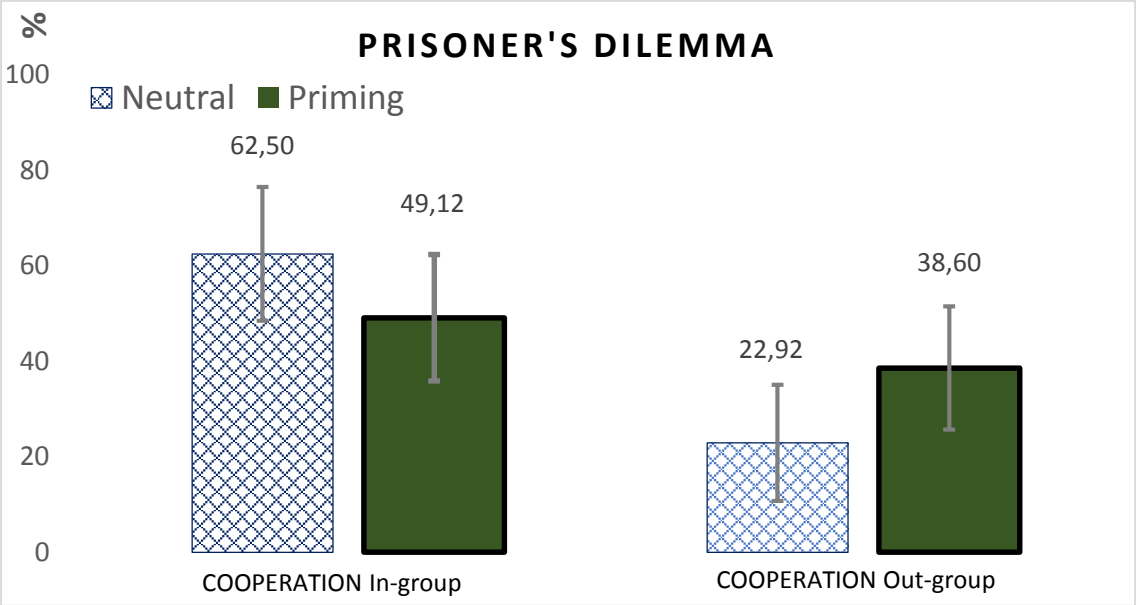


**Figure 5:** Inmates' decisions in the PD, neutral vs. priming treatment. 'Cooperation In-group' and 'Cooperation Out-group' denote cooperation rates with a partner from the in-group and out-group, respectively.

On the contrary, the priming intervention seems to have no effect on distributional choices in the non-strategic context of the EET. Figure 6 displays the distribution of types by comparing the subjects in the neutral and in the priming treatment. All comparisons are insignificant, with the partial exception of a lower fraction of inequality loving types towards the in-group ($41.7\%\ vs.\ 24.6\%;\ \chi^2 = 3.48, \mathrm{p} = 0.06, \mathrm{N} = 105$).[18]

[18] $\chi^2$ tests also confirm that the distributions of types do not differ significantly by priming condition, for choices towards the in-group or the out-group (p=0.23, p=0.88, respectively).

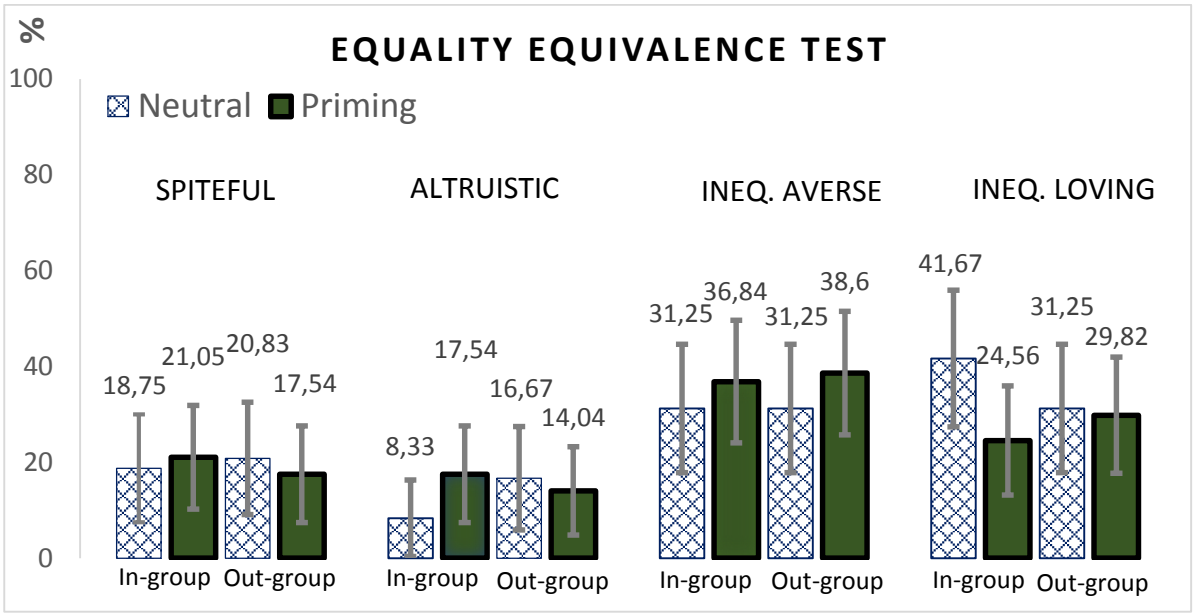


**Figure 6:** Inmates' classification of types in the EET (Spiteful, Altruistic, Inequality Averse, Inequality Loving), neutral vs. priming treatment. Bars show percentage of subjects classified into each type and include 95% confidence intervals.

*Result 2: The priming intervention increases the pro-social behavior of inmates towards the out-group by leading to higher levels of trust, reciprocity and cooperation. On the contrary, the intervention does not significantly alter the distribution of types in the EET and it has no effect on behavior towards the in-group.*

In the priming task inmates were asked to reflect about their time in prison and to think about their social behavior. Thus, the priming manipulation may have activated different concepts. One way to address this issue is to analyze the content of inmates' accounts in the priming task in order to identify the concepts activated by priming. For this purpose, we employed two independent research assistants and asked them to read all accounts in the priming treatment and to indicate for each account whether it refers to time spent in prison, social behavior, or none of the these concepts. Each assistant, independently, coded all accounts. The two codings are very highly correlated (Spearman's $\rho=0.82$, $p<0.01$, $N=57$), leading to a scale reliability coefficient (Crohnbach's alpha) of 0.89. The two assistants met and jointly decided on a common coding in all cases of disagreement. Based on this method, we classified all accounts into one of the following categories: (i) accounts referring to time spent in prison, (ii) accounts referring to social behavior, (iii) accounts referring to none of the previous two concepts.[19]

[19] A fourth possible category would include accounts referring to both, time spent in prison and social behavior. Given that this category cannot be useful in disentangling the effects of the two types of accounts, we asked our

Of 57 scripts, the majority (33 scripts, or 58%) were coded as primarily related to time spent in prison, 13 (23%) were coded as primarily related to social behavior, and 11 (19%) as related to none of the two. This already suggests that time in prison was the most prominent concept activated by priming. Moreover, in order to give an impression of how the priming effects are related to script content, in Table 4 we present behavior in the three experimental games, disaggregated by category of coded script.

**Table 4. Choices in the experimental games, by coded category of inmates' account**

| | Coded category of inmates' accounts | | |
|---|---|---|---|
| | Time in prison | Social behavior | None |
| **Trust Game**: | | | |
| Trust in-group | 36.4 | 60.0 | 80.0 |
| Trust out-group | 72.7 | 60.0 | 0.0 |
| Reciprocity in-group | 50.0 | 72.7 | 50.0 |
| Reciprocity out-group | 87.5 | 36.4 | 0.0 |
| **Prisoner's Dilemma**: | | | |
| Cooperation in-group | 46.7 | 50.0 | 54.5 |
| Cooperation outgroup | 63.3 | 18.8 | 0.0 |
| **Equality Equivalence Test**: | | | |
| Spiteful in-group | 16.7 | 25.0 | 27.3 |
| Spiteful out-group | 16.7 | 18.8 | 18.2 |
| Inequality averse in-group | 36.7 | 31.3 | 45.5 |
| Inequality averse out-group | 46.7 | 31.3 | 27.3 |
| Altruistic in-group | 16.7 | 18.8 | 18.2 |
| Altruistic out-group | 13.3 | 6.3 | 27.3 |
| Inequality loving in-group | 30.0 | 25.0 | 9.1 |
| Inequality loving out-group | 23.3 | 43.8 | 27.3 |

**Notes:** All values represent percentages. Trust measures the % of first movers who trust in the TG; Reciprocity measures the % of second movers who reciprocate in the TG; Cooperation measures the % of subjects who cooperate in the PD; Spiteful, Inequality averse, Altruistic, Inequality loving measure the % of subjects classified in each of the four categories in the EET. N=33, N=13, N=11 for the categories 'Time in prison', 'Social behavior', 'None', respectively.

assistants to decide which of the two concepts featured more prominently in the respective account for all such cases.

Table 4 suggests that the main finding with respect to the effect of priming – the increase in pro-sociality towards the out-group – is driven by those inmates who wrote about time spent in prison in their scripts. In the dimension of trust, those inmates trust the out-group 72.7% of the time, compared to only 36.4% who trust the in-group. The pattern is very similar in the dimension of reciprocity (87.5% are reciprocal towards the out-group compared with 50% towards the in-group) and cooperation in the prisoner's dilemma (63.3% cooperate with the out-group and 46.7% with the in-group). All these differences are reversed among those inmates who wrote about social behavior in their scripts, and also among those who wrote about neither time spent in prison nor social behavior (noting, however, that all numbers in these two categories are based on very few observations and should thus be interpreted with caution). Finally, consistent with the absence of treatment differences in EET classification between the primed and neutral group (see Figure 6), we find that the proportions of the four distributional preference types are typically quite similar across coded categories. Again, these figures are based on very few observations per distributional preference type and are therefore likely to be imprecise.

It is evident from the numbers shown in Table 4 that scripts referring to time spent in prison lead to very high levels of pro-sociality towards the out-group in the dimensions of trust, reciprocity and cooperation, helping us to better understand how the priming intervention worked. This analysis also hints towards the possibility that our findings actually represent a lower bound of the effects that an intervention would have if it only asked inmates to think about their time in prison.

## 5. Econometric analysis of inmates' decisions

In this section we use multivariate regression models in order to offer a richer and more nuanced analysis of the outcomes of interest. The dependent variables in Table 5 are trusting rates from the TG (columns 1-2), reciprocity rates from the TG (columns 3-4), cooperation rates in the PD (columns 5-6), as well as benevolence in the domain of advantageous and disadvantageous inequality (columns 7-10). All specifications show estimates from linear probability models. However, in Appendix A (Table A2) we also estimate Probit models (in columns 1-6) and ordered Probit models (for columns 7-10) and confirm that our results remain qualitatively unchanged. Given that each inmate made two decisions in each game (one for the in-group and one for the out-group), all specifications include individual random effects. For each dependent variable, we estimate one parsimonious specification aimed at confirming Results 1 and 2 by

means of the right-hand side dummy variables *Out-group* (equal to 1 if the matched partner is not a prisoner) and *Priming* (equal to 1 if the inmate belongs to the group that underwent the priming intervention), as well as an interaction term between the two. We also present a full specification that includes a number of potentially relevant control variables.

Among these variables, of particular interest for our research questions is *Time Served*, which is provided by the prison administration and measures (in years) the time spent by an inmate at the current prison.[20] The coefficient of this variable provides an estimate of the effect of time spent in prison on social behavior, and can thus be interpreted as an estimate of a rehabilitation effect. In order to account for the fact that a longer time served, as well as behavior in our experimental games are simultaneously determined by a lower pro-social inclination, we include in all regressions the variable *Penalty*, which measures the total sentence and, therefore, controls for the possibility that more antisocial prisoners have been longer in prison simply due to a longer initial sentence. To further account for unobservable characteristics related to pro-social inclination, we include dummy variables for the various kinds of crimes committed by the inmates (as defined in section 4.1). Robbery is the omitted category in the regressions.[21] Further control variables include the number of inmates with whom the cell is shared, as well as the inmate's age and marital status. All standard errors are clustered at the session level.

The first thing to note in the regressions is that Result 1 is fully confirmed for all games, both in the parsimonious and in the full specifications. In particular, the negative and highly significant coefficients for *Out-group* in the reciprocity and cooperation regressions confirm the presence of discrimination against the sample of non-prisoners. The magnitude of the effect of discrimination is striking: depending on specification, non-primed inmates are between 48% and 57% less likely to reciprocate towards a member of the out-group than towards another inmate, and they are between 24% and 40% less likely to cooperate with the out-group. There is also some weak evidence that inmates are less trusting towards the out-group, but this effect vanishes once the full set of controls is included.

[20] For most prisoners we also have data on the total time spent in prison over their lifetime. However, we do not use these data for our analysis since, contrary to time served in the current prison, they are self-reported and prison administration has indicated that there is no reliable way of confirming the accuracy of these data.

[21] To have an ordering of the various offences with respect to their severity, we ran simple linear regressions of the total penalty on each category independently. The offences are ordered as follows (in ascending order of severity): Robbery, smuggling goods, sexual crime, economic crime, firearms violations, drug related offences, human trafficking, homicide. Hence, the omitted category of robbery is the least severe offence. These results remain practically unchanged if we run a multivariate regression with all crime type dummies included simultaneously.

Turning to the effect of priming (Result 2), again the regression estimates are fully in line with the non-parametric analysis of section 4.3. *Priming* has an insignificant coefficient in columns 1-6, indicating that the priming intervention does not affect inmates' behavior towards other inmates with respect to trust, reciprocity, and cooperation. On the contrary, the coefficient of the interaction term *Out-group*Priming* is always positive and significant for these dimensions, confirming that priming affects behavior towards the out-group in a qualitatively different way than towards the in-group. The joint coefficients (*Priming* + *Out-group* Priming*), shown at the bottom of Table 5, capture the effect of the intervention on behavior towards the out-group, and these coefficients are positive, of large magnitude (ranging between 15% for cooperation rates and about 30%-40% for trust and reciprocity rates) and significant in five of six specifications for these three dimensions. Another way of looking at the effect of priming on discrimination is by means of the joint coefficients (*Out-group* + *Out-group*Priming*), which are smaller and generally insignificant, revealing that no systematic discrimination can be detected among primed inmates. A single exception is the positive joint coefficient in column (2), which suggests that some reverse discrimination may be present, in the form of higher trust rates towards the out-group.

With respect to distributional preferences, our regressions can provide some additional information compared to the non-parametric analysis of section 4.3. This is due to the fact that in Table 5 the dependent variable is not the individual type, but the strength of benevolence (or malevolence) towards another individual as measured by the x-score (columns 7-8) and the y-score (columns 9-10). This allows us to always use the full sample (pooling all types together) and it also allows documenting and exploiting changes in revealed benevolence that are not large enough to result in changes in type. The regression results suggest that priming affects the pro-sociality of inmates towards their in-group in two different directions: negatively in the domain of disadvantageous inequality (x-scores) and positively in the domain of advantageous inequality (y-scores). Casual inspection of the type classification in Figure 6 is in line with these effects, since both can explain the lower prevalence of inequality loving types when primed inmates play the EET with an in-group member. However, while quite sizeable, the coefficients are insignificant (with the partial exception of column 10). Given the absence of discrimination in the neutral group (see the coefficients for *Out-group* in columns 7-10) and the insignificant impact of priming on inmates' behavior, it is no surprise that we also fail to find any discrimination against the out-group among primed inmates (joint coefficients *Out-group+ Out-group* Priming*).

In addition to providing further support for our main results, the regressions of Table 5 allow us to estimate in the five full specifications the effect of spending time in prison (controlling for the total sentence), as measured by the variable *Time Served*. The coefficient of this variable is generally insignificant with the exception of specification (10), suggesting that a longer time spent in prison reduces benevolence when lying ahead of another participant but does not affect behavior in the other domains. However, this coefficient captures the effect of time served on behavior towards the in-group: the picture changes if we examine the joint coefficient *(Time Served + Out-group*Time Served)*, which is again significantly negative in the y-score regression of column (10), but in addition has a significant negative coefficient in the trust regression in (2) as well as in the cooperation regression in (4).[22] Taken together, these findings suggest that a longer time spent in prison generally reduces pro-social behavior towards the out-group. From a policy perspective this is quite worrying, since it suggests that pro-social motivation among inmates seems to wane over time, possibly rendering their re-integration after release particularly problematic. Our priming intervention is offering a way to counter-balance this negative effect of time spent in prison, given that it has already been shown to increase pro-sociality, especially towards the out-group.

*Result 3: A longer time spent in prison is associated with lower trust and cooperation rates towards the out-group, as well as decreased benevolence towards both groups when lying ahead of another participant.*

While the independent variables *Out-group* and *Priming* that have been used to analyze discrimination and the effects of priming are by design exogenous as part of the experiment, a potential concern with *Time Served* is that it may suffer from a simultaneity bias that would impede causal inference on how time spent in prison affects pro-social behavior in the Table 5 regressions. Hence, although the regressions include a number of variables to account for this possibility, causality cannot be established with certainty and these results must be interpreted with caution. As an additional step in this respect, we have also estimated two-stage least squares versions of the five full specifications. In those regressions we use the following two instruments for *Time Served*: the frequency of leaving prison (measured in days per month) and the number of days per month doing some work in the facility. Both variables are likely to

[22] We have also estimated versions of all regressions that include an interaction term between time served and priming, as well as the three-way interaction term *Out-group*Priming*Time Served*. The interaction between time served and priming is always insignificant with one exception (the cooperation regression), while the three-way interaction term is always insignificant and small in magnitude. The results are available from the authors upon request.

correlate with the time spent in prison (for instance, because they depend on the total sentence received or because certain privileges are only awarded after completing part of the sentence), while they are unlikely determinants of behavior in our experimental games.[23] In terms of instrument strength, we note that working days per month – but not the frequency of leaving prison – is a significant predictor of time served in all first stage regressions, and that testing for instrument strength based on first stage F statistics (following Stock and Yogo, 2005) rejects the null hypothesis of weak instruments in all but one (the reciprocity) specification.

Our main results remain robust to instrumental variables estimation, painting a consistent picture with respect to the negative effects of time spent in prison on pro-social behavior, primarily towards the out-group: the significant negative effects of *Time Served* persist for the y-score (both towards the in-group and the out-group), as well as in the dimensions of trust and cooperation towards the out-group. However, we relegate this analysis to Table A3 in Appendix A, due to the fact that Wu-Hausman tests of endogeneity fail to reject the null hypothesis that *Time Served* can be treated as exogenous, albeit with one exception in the regression on y-scores (see bottom of Table A3 for details on tests of endogeneity and of over-identifying restrictions).

In terms of further control variables (*Penalty, High Security, Number of Cell Mates*, *Age*, socioeconomic controls and the crime categories), we note that out of 75 estimated coefficients for these variables in columns (2), (4), (6), (8) and (10) of Table 5, only 16 are significant at least at the 10% level, and only the variable *Smuggling Goods* is significant in more than two specifications. For instance, the most severe crime of homicide is associated with significantly lower reciprocity levels but has positive and insignificant coefficients in all other specifications. Overall, it is hard to discern any systematic pattern regarding the pro-sociality attributed to different types of committed crimes. Inmates from the high security prison are, ceteris paribus, less reciprocal, but there are no significant differences between the two prisons in any other dimension – which may, however, also be due to the small sample in the high security prison.

[23] We must, however, acknowledge that this analysis must be interpreted with caution since instrument exogeneity cannot be tested. Sargan's tests of overidentifying restrictions, reported in Table A3, do not reject the null hypothesis that all instruments are exogenous, under the assumption that at least one of them is exogenous.

**Table 5: Econometric analysis**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Trust** | **Trust** | **Recipr.** | **Recipr.** | **Cooper.** | **Cooper.** | **x-score** | **x-score** | **y-score** | **y-score** |
| *Out-group* | -0.222* | -0.112 | -0.571*** | -0.483*** | -0.396*** | -0.238** | -0.208 | 0.021 | 0.208 | 0.006 |
| | (0.125) | (0.163) | (0.134) | (0.163) | (0.093) | (0.114) | (0.345) | (0.439) | (0.328) | (0.406) |
| *Priming* | 0.061 | -0.031 | -0.114 | -0.050 | -0.134 | -0.153 | -0.485 | -0.533 | 0.612 | 0.899* |
| | (0.125) | (0.142) | (0.138) | (0.158) | (0.094) | (0.103) | (0.419) | (0.471) | (0.447) | (0.471) |
| *Out-group*Priming* | 0.347** | 0.380** | 0.411** | 0.343* | 0.291** | 0.305** | 0.542 | 0.410 | -0.226 | -0.187 |
| | (0.170) | (0.183) | (0.181) | (0.201) | (0.126) | (0.133) | (0.468) | (0.512) | (0.446) | (0.474) |
| *Time Served* | | -0.018 | | 0.021 | | 0.001 | | 0.035 | | -0.198*** |
| | | (0.019) | | (0.023) | | (0.014) | | (0.063) | | (0.063) |
| *Out-group*Time Served* | | -0.024 | | -0.008 | | -0.037** | | -0.028 | | 0.013 |
| | | (0.023) | | (0.028) | | (0.018) | | (0.068) | | (0.063) |
| *Penalty* | | -0.007 | | 0.020** | | -0.004 | | 0.006 | | -0.029 |
| | | (0.009) | | (0.008) | | (0.006) | | (0.026) | | (0.027) |
| *High Security* | | -0.167 | | -0.721** | | 0.133 | | -0.294 | | 1.423 |
| | | (0.308) | | (0.332) | | (0.205) | | (1.026) | | (1.059) |
| *Married* | | 0.045 | | -0.184 | | 0.147 | | 0.289 | | -0.154 |
| | | (0.126) | | (0.192) | | (0.091) | | (0.457) | | (0.471) |
| *Number of Cell Mates* | | -0.099** | | -0.026 | | -0.030 | | 0.120 | | 0.141 |
| | | (0.039) | | (0.048) | | (0.027) | | (0.135) | | (0.139) |
| *Age* | | 0.002 | | -0.011 | | -0.007 | | 0.052** | | -0.015 |
| | | (0.008) | | (0.008) | | (0.005) | | (0.026) | | (0.026) |
| *Education* | | -0.045 | | -0.103 | | -0.005 | | 0.012 | | -0.400* |
| | | (0.062) | | (0.087) | | (0.046) | | (0.228) | | (0.235) |
| *Number of Children* | | 0.043 | | 0.073 | | 0.020 | | -0.282* | | -0.090 |
| | | (0.037) | | (0.083) | | (0.033) | | (0.160) | | (0.165) |
| *Number of Siblings* | | 0.019* | | -0.029 | | -0.009 | | 0.050 | | -0.030 |
| | | (0.011) | | (0.022) | | (0.009) | | (0.046) | | (0.047) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Drug Related Offences* | | -0.167 | | 0.055 | | 0.210** | | -0.241 | | 0.684 |
| | | (0.117) | | (0.135) | | (0.084) | | (0.421) | | (0.435) |
| *Smuggling Goods* | | 0.805* | | 0.351* | | -0.046 | | -0.906 | | 1.573* |
| | | (0.428) | | (0.211) | | (0.161) | | (0.807) | | (0.832) |
| *Human Trafficking* | | -0.071 | | -0.128 | | 0.220* | | -0.602 | | 0.265 |
| | | (0.157) | | (0.256) | | (0.121) | | (0.608) | | (0.627) |
| *Homicide* | | 0.825 | | -0.839** | | 0.315 | | 0.714 | | 1.569 |
| | | (0.532) | | (0.422) | | (0.318) | | (1.595) | | (1.645) |
| *Sexual Crime* | | -0.156 | | 0.261 | | -0.021 | | 1.092 | | 0.153 |
| | | (0.277) | | (0.433) | | (0.214) | | (1.074) | | (1.108) |
| *Economic Crime* | | 0.195 | | 0.105 | | 0.075 | | -1.475** | | 0.496 |
| | | (0.185) | | (0.263) | | (0.139) | | (0.698) | | (0.721) |
| *Firearms violations* | | 0.305 | | -0.308 | | -0.020 | | -1.436* | | 1.025 |
| | | (0.202) | | (0.405) | | (0.167) | | (0.836) | | (0.863) |
| *N* | 118 | 102 | 92 | 84 | 210 | 186 | 210 | 186 | 210 | 186 |
| *Priming + Out-group* Priming* | 0.408*** | 0.349** | 0.297** | 0.293* | 0.157* | 0.152 | 0.057 | -0.123 | 0.386 | 0.712 |
| *Out-group + Out-group* Priming* | 0.125 | 0.268* | -0.160 | -0.140 | -0.105 | 0.067 | 0.334 | 0.431 | -0.018 | -0.181 |
| *Time Served + Out-group*Time Served* | n/a | -0.042** | n/a | 0.013 | n/a | -0.037*** | n/a | 0.006 | n/a | -0.185*** |

**Notes:** All specifications are estimated with linear probability models and include subject random effects. Standard errors in parantheses. Independent variables as defined in the text and in footnote 16. Last three rows present joint coefficients and significance levels for tests on the restriction that the respective joint coefficient is equal to zero. N varies across specifications depending on the number of available observations (59 inmates as first movers, 46 inmates as second movers in the TG; data on control variables for some inmates missing in full specifications). *, **, *** indicates statistical significance at the 10%, 5%, 1% level, respectively.

## 6. Discussion and conclusion

This paper presents evidence on the social behavior of prison inmates, with the aim of characterizing this behavior and also of evaluating a priming intervention and its potential to promote a positive rehabilitation effect. We have run economic experiments with a sample of 105 inmates in a low- and a high security prison in Greece. We have provided a general landscape of inmates' social behavior and studied how this varies between the in-group of inmates and the out-group of people outside prison, finding substantial evidence of discrimination expressed in lower rates of reciprocity and cooperation towards the out-group. Furthermore, we have used a simple priming technique that asks inmates to reflect on their time spent in prison and how it has affected them: the results indicate that this priming intervention has a substantial impact on inmates' choices, generally promoting pro-social behavior towards the out-group and eliminating discrimination.

We believe that our findings can be very useful to policy makers who are in charge of evaluating the impact of incarceration on social behavior and the likelihood of successful post-release re-integration, especially in light of the controversy surrounding the way that prisons affect inmates during and after incarceration. While our results generally point towards a negative effect of time spent in prison on pro-social behavior (evinced by lower cooperation rates when interacting with the out-group and by lower y-scores), the strong positive effect of our priming intervention suggests that rehabilitation can be promoted with help of the appropriate instruments. We have evaluated one such instrument and documented some very encouraging results – which are, however, subject to a number of limitations discussed in what follows.

What are possible channels driving the positive effect of the priming intervention in our study? The reasoning behind our hypothesis H2 proposes a mechanism based on concerns for group reputation, which are stronger in the priming treatment that enhances the participants' identification with the prisoner group. We have already shown that the prisoner's identity significantly increases in strength as a result of priming, which is consistent with our story. However, further possible mechanisms for the effect of priming should be acknowledged. First, it seems plausible that emotions evoked by the priming task may be driving the effects. For instance, the pattern we observe in the data (stronger identification with the prisoners' group and stronger pro-sociality towards the out-group in the priming treatment) seems quite reasonable if identification with the in-group is associated with generally negative feelings and experiences, such that inmates who are asked to reflect about their time in prison regret sharing

this particular identity and become more inclined to behave benevolently towards the group from outside prison (for instance, in an attempt to shed this criminal identity, or to partly compensate for previous criminal behavior). Referring to the content analysis presented in section 3, this interpretation is in line with the much higher frequency of primed inmates' accounts of their time in prison being classified as generally negative, compared to the neutral group accounts (50% vs. 8.5%; $\chi^2 = 20.33, p < 0.01, N = 101$).

In a similar vein, the positive effect of priming on pro-social behavior towards the out-group may be associated with subordination towards the depriving out-group (Giles and Powesland, 1975). If inmates consider society as a depriving out-group, then they may develop a stronger inclination to behave benevolently towards this group when their prisoner's identity is made more salient. This explanation is supported by the presence of reverse discrimination among primed inmates in the dimensions of trust and benevolence under disadvantageous inequality.

A final possibility is that the priming intervention evokes among inmates the rehabilitation effect that it aims to test. For instance, introspection may lead inmates to consider the purpose that their incarceration is meant to serve, and to realize that imprisonment can only be meaningful if it can lead to successful re-integration by modifying pro- and antisocial behavior. A related interpretation suggests that the effects of priming are driven by a social desirability bias: if inmates understand that the intervention is meant to test their rehabilitation, they may adjust their choices to fit the expected pattern of successful rehabilitation and increased pro-sociality towards the out-of-prison sample. Regardless of the precise mechanism driving our findings, the fact that the intervention works is very encouraging and suggests that perhaps further, related interventions, may have the capacity to modify inmates' behavior.

Given that participation in the experiments was voluntary, issues of selection into the sample may potentially play a role. In order to alleviate such concerns and assess the extent to which our sample is representative of the general prison population, we compare certain observable characteristics in our sample and among all prisoners in Greece and in the United States. The first characteristic is the distribution of nationalities, differentiating between Greeks and foreigners. For this analysis we use data from the Greek Ministry of Justice, Transparency and Human Rights[24] and find that in our sample, as well as in the prison population in Greece as a whole, foreigners are a majority (70% and 58%, respectively). The second characteristic is

[24] Available at http://www.ministryofjustice.gr/site/en/PenitentiarySystem/Statisticaldataondetainees.aspx; accessed November 2017.

the type of crime committed by inmates, for which we have no suitable data for Greek prisons and therefore use U.S. data from the Federal Bureau of Prisons.[25] Table A4 in Appendix A shows that the distribution of crime categories is similar between our sample and the U.S. aggregate data, although it should be taken into account that the definitions of various categories do not completely overlap. Finally, from the same source we have data on the mean age of prisoners in the U.S. (38.4 years), which is very similar to the average age of 36 years in our sample.

The experimental methods used in this study inevitably suffer from certain limitations, thus only allowing us to draw indirect evidence on our research questions. The data are based on behavior in simple economic games, which is only a proxy for criminal or pro-social behavior in the field. Moreover, behavior is recorded before release from prison, meaning that the evidence on how inmates treat non-inmates refers to a point in time prior to actual encounters between the two groups. It should also be kept in mind that, given imperfections in the legal system, it cannot be ruled out that some of the inmates in our sample have been wrongly convicted.

An additional limitation of our study is that it paints a static picture of inmates' behavior, documenting their behavior at one point in time and using econometric techniques in order to estimate an effect of time spent in prison on behavior. Relatedly, we cannot be sure about the extent to which the strong positive effects of priming documented in our experiment persist in the long run, given that priming effects are often short-lived. An alternative method could thus be to identify the effect of imprisonment and of priming interventions by following inmates over time and documenting changes in their behavior during their sentence as well as after their release from prison. While this can be challenging due to issues of attrition over time and biases resulting from repeated measurement of social preferences, we consider it a very promising avenue for future research.

***Acknowledgements*:** We are thankful to Alexander Cappelen, Yan Chen, Alain Cohn, Christoph Engel, Marie-Claire Villeval, participants at IMEBESS 2017, ESA Asian Pacific Meetings 2017, the Austrian Experimental Workshop 2017, Bonn-Innsbruck-Munich workshop, and seminars in Alicante, Berlin, Birmingham, Cologne, Exeter, Innsbruck, Marburg, Potsdam, Siena, Swansea and UA Madrid, for valuable comments. Special thanks

[25] Available at https://www.bop.gov/about/statistics/statistics_inmate_offenses.jsp; accessed November 2017.

are due to the Hellenic Ministry of Justice, Transparency and Human Rights, administration and staff at the high security prison facility “Crete 1” and the agricultural prison facility of “Agia”. We thank Alexandra Baier and Sarah Stuefer for research assistance. We thankfully acknowledge financial support from the British Academy (BA LEVER 2) and the Spanish Ministerio de Economía y Competitividad (grant ECO2015-68469-R AEI/FEDER). This project has received IRB approval from the Ethics Committee of the School of Agriculture Policy and Development at the University of Reading.

# References


Akerlof, George A. 1982. “Labor Contracts as Partial Gift Exchange. ” *The Quarterly Journal of Economics* 97 (4): 543–69.

Anderson, David A. 2012. “The Cost of Crime.” *Foundations and Trends in Microeconomics* 7 (3): 209–65.

Andersson, Ola, Topi Miettinen, Kaisa Hytönen, Magnus Johannesson, and Ute Stephan. 2017. “Subliminal Influence on Generosity.” *Experimental Economics* 20 (3): 531–55.

Balafoutas, Loukas, Rudolf Kerschbamer, Martin Kocher, and Matthias Sutter. 2014. “Revealed Distributional Preferences: Individuals vs. Teams.” *Journal of Economic Behavior and Organization* 108: 319–30.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter. 2012. “Distributional Preferences and Competitive Behavior.” *Journal of Economic Behavior and Organization* 83 (1): 125–35.

Bargh, John A., and Tanya L. Chartrand. 2000. “Studying the Mind in the Middle: A Practical Guide to Priming and Automaticity Research.” In *Handbook of Research Methods in Social and Personality Psychology*, edited by C.M. Judd and H.T. Reis, 253–85.

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts. 2006. “Cues of Being Watched Enhance Cooperation in a Real-World Setting.” *Biology Letters* 2 (3): 412–14.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. “Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections.” *The Quarterly Journal of Economics* 124 (1): 105–47.

Benjamin, Daniel J., James J. Choi, and Geoffrey Fisher. 2016. “Religious Identity and Economic Behavior.” *The Review of Economics and Statistics* 98 (4): 617–37.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. “Trust, Reciprocity, and Social History.” *Games and Economic Behavior* 10 (1): 122–42.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2016. “Incarceration, Recidivism and Employment.” *NBER Working Paper 22648*.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. 2018. “Incarceration Spillovers in Criminal and Family Networks.” *NBER Working Paper 24878.*

Birkeland, Sigbjørn, Alexander W. Cappelen, Erik O. Sørensen, and Bertil Tungodden. 2014. “An Experimental Study of Prosocial Motivation among Criminals.” *Experimental Economics* 17 (4): 501–11.

Bucciol, Alessandro, and Marco Piovesan. 2011. “Luck or Cheating? A Field Experiment on Honesty with Children.” *Journal of Economic Psychology* 32: 73–78.

Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger. 2014. “Violence and Risk Preference: Experimental Evidence from Afghanistan.” *The American Economic Review* 104 (1): 123–48.

Charness, Gary, and Chun-Lei Yang. 2014. “Starting Small toward Voluntary Formation of Efficient Large Groups in Public Goods Provision.” *Journal of Economic Behavior and Organization* 102: 119–32.

Chen, M. Keith, and Jesse M. Shapiro. 2007. “Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-based Approach.” *American Law and Economics Review* 9(1): 1–29.

Chen, Yan, and Sherry Xin Li. 2009. “Group Identity and Social Preferences.” *American Economic Review* 99 (1): 431–57.

Chmura, Thorsten, Christoph Engel, and Markus Englerth. 2016. “At the Mercy of a Prisoner Three Dictator Experiments.” *Applied Economics Letters* 24 (11): 1–5.

Chmura, Thorsten, Christoph Engel, Markus Englerth, and Thomas Pitz. 2010. “At the Mercy of the Prisoner Next Door - Using an Experimental Measure of Selfishness as a Criminological Tool.” *Max Planck Institute Working Papers Bonn* 27: 1–24.

Cinyabuguma, Matthias, Talbot Page, and Louis Putterman. 2005. “Cooperation under the Threat of Expulsion in a Public Goods Experiment.” *Journal of Public Economics* 89 (8): 1421–35.

Clark, Brendan C., Christopher B. Thorne, Peter S. Hendricks, Carla Sharp, Shane K. Clark, and Karen L. Cropsey. 2015. “Individuals in the Criminal Justice System Show Differences in Cooperative Behaviour: Implications from Cooperative Games.” *Criminal Behaviour and Mental Health* 21 (5): 299–306.

Cohn, Alain, Jan Engelmann, Ernst Fehr, and Michel André Maréchal. 2015. “Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals.” *American Economic Review* 105 (2): 860–85.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal. 2014. “Business Culture and Dishonesty in the Banking Industry.” *Nature* 516 (7529): 86–89.

Cohn, Alain, and Michel André Maréchal. 2016. “Priming in Economics.” *Current Opinions in Psychology* 12: 17–21.

Cohn, Alain, Michel André Maréchal, and Thomas Noll. 2015. “Bad Boys: How Criminal Identity Salience Affects Rule Violation.” *Review of Economic Studies* 82 (4): 1289–1308.

Cox, James C. 2004. “How to Identify Trust and Reciprocity.” *Games and Economic Behavior* 46 (2): 260–81.

Drago, Francesco, Roberto Galbiati, and Pietro Vertova. 2011. “Prison Conditions and Recidivism.” *American Law and Economics Review* 13 (1): 103–30.

Durose, Matthew R, Alexia Cooper, and Howard Snyder. 2014. “Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010.” *U.S. Department of Justice* Special Re (April): 1–31.

Ekström, Mathias. 2012. “Do Watching Eyes Affect Charitable Giving? Evidence from a Field Experiment.” *Experimental Economics* 15 (3): 530–46.

Everett, Jim A.C., Nadira S. Faber, and Molly Crockett. 2015. “Preferences and Beliefs in In-group Favoritism.” *Frontiers in Behavioral Neuroscience* 9:15.

Farrington, David P., and Brandon C. Welsh. 2005. “Randomized Experiments in Criminology: What Have We Learned in theLlast Two Decades?” *Journal of Experimental Criminology* 1 (1): 9–38.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. “Reciprocity as a Contract Enforcement Device: Experimental Evidence.” *Econometrica* 65: 833–60.

Fehr, Ernst, and Frédéric Schneider. 2010. “Eyes Are on Us, but Nobody Cares: Are Eye Cues Relevant for Strong Reciprocity?” *Proceedings of the Royal Society B: Biological Sciences* 277 (1686): 1315–23.

Fehr, Ernst, and Andreas Leibbrandt. 2011. “A Field Study on Cooperativeness and Impatience in the Tragedy of the Commons.” *Journal of Public Economics* 95 (9-10): 1144–55.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. 2017. “Distributional Preferences and Political Behavior.” *Journal of Public Economics* 155: 1–10.

Gendreau, Paul, Claire Goggin, and Francis T. Cullen. 1999. “The Effects of Prison Sentences on Recidivism.” *Public Works and Government Services Canada*.

Giles, Howard, and Peter F. Powesland. 1975. *Speech Style and Social Evaluation*. Oxford, England: Academic Press.

Gino, Francesca, Shahar Ayal, and Dan Ariely. 2009. “Contagion and Differentiation in Unethical Behavior.” *Psychological Science* 20 (3): 393–98.

Gneezy, Ury, Andreas Leibbrandt, and John A. List. 2016. “Ode to the Sea: Workplace Organizations and Norms of Cooperation.” *The Economic Journal* 126 (595): 1856-83.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. “The Impact of Group Membership on Cooperation and Norm Enforcement.” *American Economic Review* 96 (2): 212–16.

Haley, Kevin J., and Daniel M.T. Fessler. 2005. “Nobody’s Watching?” *Evolution and Human*

*Behavior* 26 (3): 245–56.

Harrison, Glenn W., and John List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–55.

Hopkins, Nick, Reicher, Steve, Harrison, Kate, Cassidy, Clare, Bull, Rebecca, and Mark Levine. 2007. "Helping to Improve the Group Stereotype: On the Strategic Dimension of Prosocial Behavior." *Personality and Social Psychology Bulletin* 33, 776–788.

Isoni, Andrea, and Robert Sugden. 2018. "Reciprocity and the Paradox of Trust in psychological game theory. " *Journal of Economic Behavior and Organization,* in press.

Kerr, Norbert L., Ann C. Rumble, Ernest S. Park, Jaap W. Ouwerkerk, Craig D. Parks, Marcello Gallucci, and Paul A. M. Van Lange. 2009. "How Many Bad Apples Does it Take to Spoil the Whole Barrel?: Social Exclusion and Toleration for Bad Apples." *Journal of Experimental Social Psychology* 45 (4): 603–13.

Kerschbamer, Rudolf. 2015. "The Geometry of Distributional Preferences and a Non-Parametric Identification Approach: The Equality Equivalence Test." *European Economic Review* 76: 85–103.

Kerschbamer, Rudolf, Matthias Sutter, and Uwe Dulleck. 2017. "How Social Preferences Shape Incentives in (Experimental) Markets for Credence Goods." *Economic Journal* 127 (600): 393–416.

Khadjavi, Menusch. 2015. "Deterrence Works for Criminals." *European Journal of Law and Economics* 31 (2): 1–14.

Khadjavi, Menusch, and Andreas Lange. 2013. "Prisoners and Their Dilemma." *Journal of Economic Behavior and Organization* 92: 163–75.

Killias, Martin, Marcelo Aebi, and Denis Ribeaud. 2000. "Does Community Service Rehabilitate better than Short-term Imprisonment?: Results of a Controlled Experiment." *The Howard Journal of Criminal Justice* 39(1), 40–57.

Killias, Martin, Gwladys Gilliéron, Françoise Villard, and Clara Poglia. 2010. "How Damaging is Imprisonment in the Long-term? A Controlled Experiment Comparing Long-term Effects of Community Service and Short Custodial Sentences on Re-offending and Social Integration." *Journal of Experimental Criminology* 6(2): 115–130.

Lipsey, Mark W., and Francis T. Cullen. 2007. "The Effectiveness of Correctional Rehabilitation: A Review of Systematic Reviews." *Annual Review of Law and Social Sciences* 3: 297–320.

Maggioni, Mario A., Domenico Rossignoli, Simona Beretta, and Sara Balestri. 2018. "Trust behind Bars: Measuring Change in Inmates' Prosocial Preferences." *Journal of Economic*

*Psychology* 64: 89–104.

Maier-Rigaud, Frank P., Peter Martinsson, and Gianandrea Staffiero. 2010. "Ostracism and the Provision of a Public Good: Experimental Evidence." *Journal of Economic Behavior and Organization* 73 (3): 387–95.

Masclet, David. 2003. "Ostracism in Work Teams: A Public Good Experiment." *International Journal of Manpower* 24 (7): 867–87.

Needels, Karen E. 1996. "Go Directly to Jail and Do Not Collect? A Long-Term Study of Recidivism, Employment, and Earnings Patterns among Prison Releasees." *Journal of Research in Crime and Delinquency* 33 (4): 471–96.

Nese, Annamaria, Niall O Higgins, Patrizia Sbriglia, and Maurizio Scudiero. 2018. "Cooperation, Punishment and Organized Crime: A Lab in the Field Experiment in Southern Italy." *European Economic Review* 107: 86–98.

Phelan, Christopher, and Aldo Rustichini. 2018. "Pareto Efficiency and Identity." *Theoretical Economics* 13: 979–1008.

Putterman, Louis. 2006. "Reciprocity, Altruism, and Cooperative Production." *Handbook of the Economics of Giving, Altruism and Reciprocity* 2: 1409–35.

Romano, Angelo, Daniel Balliet, Toshio Yamagishi, and James H. Liu. 2017. "Parochial Trust and Cooperation across 17 Societies." *Proceedings of the National Academy of Sciences* 114 (48): 12702–7.

Roodman, David. 2017. "The Impacts of Incarceration on Crime." *Open Philantropy Project* 1–142.

Rotemberg, Julio J. 2006. "Altruism, Reciprocity and Cooperation in the Workplace." *Handbook of the Economics of Giving, Altruism and Reciprocity* 2: 1371–1407.

Rusch, Hannes. 2014. "The Evolutionary Interplay of Intergroup Conflict and Altruism in Humans: A Review of Parochial Altruism Theory and Prospects for its Extension." *Proceedings of Biological Sciences B* 281: 1539.

Shariff, Azim F., and Ara Norenzayan. 2007. "God Is Watching You." *Psychological Science* 18 (9): 803–9.

Sheremeta, Roman M., Steve James Tucker, and Jingjing Zhang. 2011. "Creating Self-Sustained Social Norms through Communication and Ostracism." *19th International Congress on Modelling and Simulation*, 12–16.

Solda, Alice, and Marie-Claire Villeval. 2017. "Exclusion and Reintegration in a Social Dilemma." *GATE Lyon Saint-Etienne Working Paper 1720*.

Stock, J. H., and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." *In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. D. W. K. Andrews and J. H. Stock,* 80–108.

Tajfel, Henri. 1974. Social Identity and Intergroup Behaviour. *Social Science Information* 13: 65–93.

Toch, Hans. 1977. *Living in Prison: The Ecology of Survival*. New York: Free Press.

Villettaz, Patrice, Gwladys Gilliéron, and Martin Killias. 2014. "The Effects on Re-offending of Custodial versus Non-custodial Sanction." *National Council from Crime Prevention* 1-104.

Vohs, Kathleen D., Nicole L. Mead, and Miranda R. Goode. 2006. "The Psychological Consequences of Money." *Science* 314 (5802): 1154–56.

Walmsley, Roy. 2016. "World Prison Population List - eleventh edition." *World Prison Brief - Institute for Criminal Policy Research*.

Zizzo, Daniel John. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics* 13 (1): 75–98.

**APPENDIX A**

# 1. Comparison between inmates and students

We begin by noting that we observe some similarities between the two samples, since around 30% of participants have a Greek nationality and the rest have a variety of nationalities dominated by Arab and Eastern European countries (proportion of Greek prisoners: 29.5% in prisons vs. 24.5% in MAICH; $\chi^2 = 0.06, \mathrm{p} = 0.81, \mathrm{N} = 145$). On the other hand, a difference is that participants are on average older among the inmate population: mean age in the student group is 26.77 (s.d. = 5.18) and in the group of inmates it is 35.97 (s.d.=9.47). In any case, we are presenting evidence from two different cohorts of the population, not only in terms of certain sociodemographic features, but also due to obvious differences in life experiences and in particular the current social and professional situation. Thus, our rationale when comparing the two samples is not to isolate or causally estimate some sort of 'prison' effect; the purpose of this exercise is solely to show how the behavior of our focus group of prison inmates compares to a type of population typically studied in economic experiments (a student sample), and sharing a few common features with the prison population (close interactions among participants who eat, sleep and work together in the study site; similar distribution of nationalities; geographic proximity).

Figure A1 shows the results from the trust game. Overall, inmates trust significantly less than students (42.4% vs. 60%; $\chi^2 = 3.73, \mathrm{p} = 0.05, \mathrm{N} = 158$). They also exhibit somewhat less reciprocal behavior as second movers, but the difference is not significant ($\chi^2 = 2.41, \mathrm{p} = 0.12, \mathrm{N} = 132$).

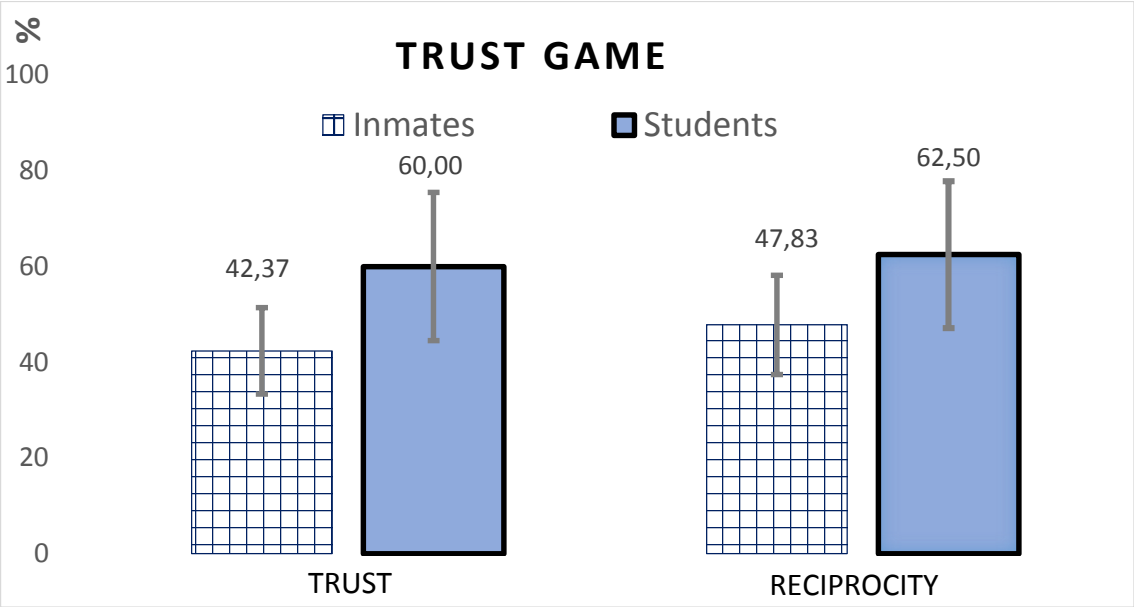


**Figure A1:** Inmates' and students' decisions in the TG. Trust is measured as percentage of first movers who choose to trust in the TG; reciprocity is measured as percentage of second movers who choose to reciprocate in the TG. Bars include 95% confidence intervals.

Figure A2 shows the results from the prisoner's dilemma. The frequency of cooperative choices is very similar across the two samples; 43.33% of inmates and 40% of our student group cooperate. These proportions are statistically indistinguishable from each other ($\chi^2 = 0.26, \mathrm{p} = 0.61, \mathrm{N} = 290$).

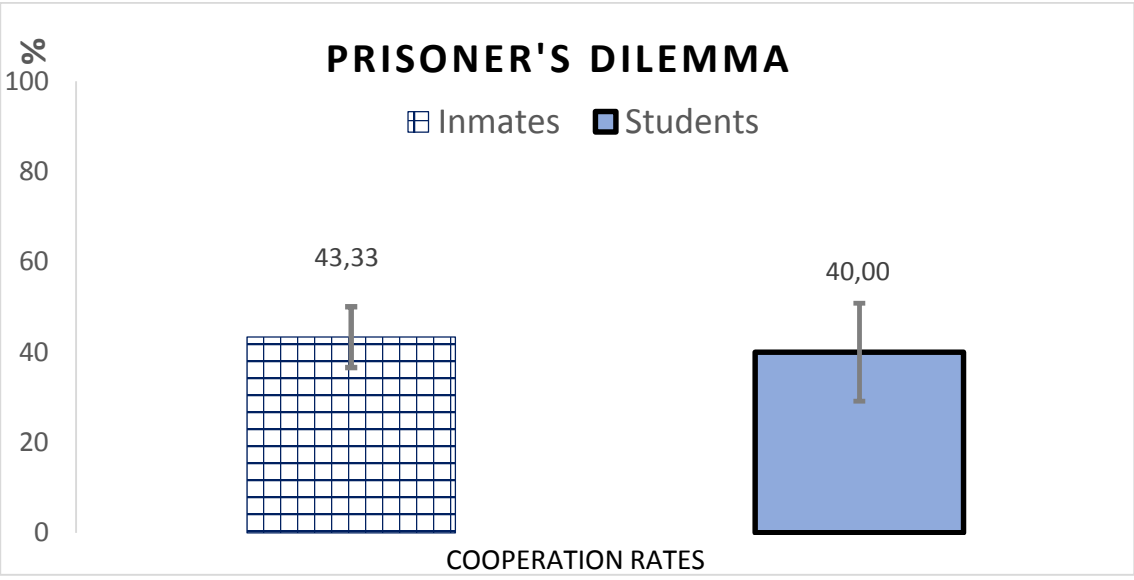


**Figure A2:** Inmates' and students' decisions in the PD. Bars show cooperation rates and include 95% confidence intervals.

Finally, Figure A3 gives clear evidence that the distribution of distributional preference types differ between the two samples. While approximately 20% of inmates and students are classified as spiteful ($\chi^2 = 0.08, \mathrm{p} = 0.93, \mathrm{N} = 290$), inmates are about half as likely as students to be classified as altruistic (14.3% vs. 28.8%; $\chi^2 = 6.61, \mathrm{p} = 0.01, \mathrm{N} = 290$), indicating lower benevolence. Similarly, the ratio of inequality averse subjects is slightly significantly higher among students than among inmates ($\chi^2 = 3.25, \mathrm{p} = 0.07, \mathrm{N} = 290$). As already noted, a striking feature of the distribution of types among prisoners is the very high prevalence of inequality loving types, which is more than six times higher than among students ($\chi^2 = 22.09, \mathrm{p} = 0.01, \mathrm{N} = 290$),

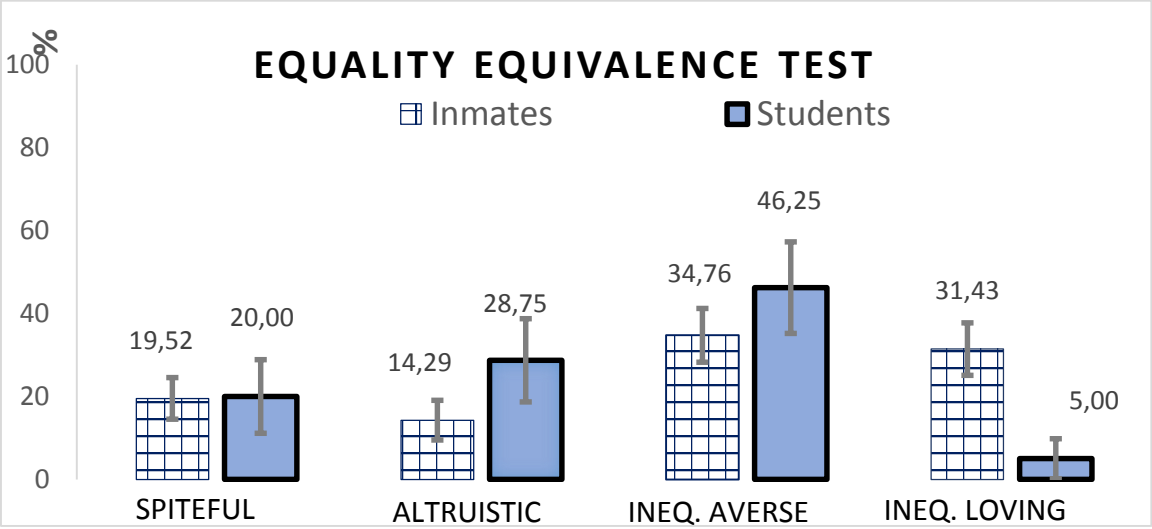


**Figure A3:** Inmates' and students' classification of types (Spiteful, Altruistic, Inequality Averse, Inequality Loving) in the EET. Bars show the percentage of subjects classified into each type and include 95% confidence intervals.

The following result summarizes the observed differences in the social behavior of inmates compared to the group of students.

*Result A1: Inmates trust less and are less altruistic, less inequality averse and more inequality loving than students. No significant differences between the two samples are found in cooperation and reciprocity rates.*

# 2. Additional tables

**Table A1: Distribution of language of instructions and of the priming text**

| Language | N | % | N | % | N | % |
|---|---|---|---|---|---|---|
| **Instructions** | | | | | | |
| ARABIC | 20 | 19.05 | 7 | 14.58 | 13 | 22.81 |
| ENGLISH | 6 | 5.71 | 3 | 6.25 | 3 | 5.26 |
| FRENCH | 2 | 1.90 | 2 | 4.17 | 0 | 0.00 |
| GREEK | 77 | 73.33 | 36 | 75.00 | 41 | 71.93 |
| **Priming text** | | | | | | |
| ARABIC | 26 | 24.76 | 12 | 25.00 | 14 | 24.56 |
| ENGLISH | 6 | 5.71 | 4 | 8.33 | 2 | 3.51 |
| GREEK | 65 | 61.90 | 29 | 60.42 | 36 | 63.16 |
| SPANISH | 5 | 4.76 | 1 | 2.08 | 4 | 7.02 |
| TURKISH | 1 | 0.95 | 0 | 0.00 | 1 | 1.75 |
| URDU | 2 | 1.90 | 2 | 4.17 | 0 | 0.00 |

**Table A2: Econometric analysis with discrete choice models**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Trust** | **Trust** | **Recipr.** | **Recipr.** | **Cooper.** | **Cooper.** | **x-score** | **x-score** | **y-score** | **y-score** |
| *Out-group* | -0.707* | -0.351 | -1.823*** | -1.574*** | -1.162*** | -0.705** | -0.132 | -0.005 | 0.153 | 0.021 |
| | (0.391) | (0.502) | (0.537) | (0.552) | (0.304) | (0.345) | (0.246) | (0.312) | (0.274) | (0.335) |
| *Priming* | 0.166 | -0.140 | -0.345 | -0.158 | -0.373 | -0.448 | -0.338 | -0.344 | 0.506 | 0.718** |
| | (0.354) | (0.428) | (0.435) | (0.479) | (0.274) | (0.288) | (0.307) | (0.319) | (0.372) | (0.366) |
| *Out-group*Priming* | 1.047** | 1.299** | 1.365** | 1.103 | 0.869** | 0.968** | 0.404 | 0.318 | -0.143 | -0.162 |
| | (0.516) | (0.585) | (0.631) | (0.673) | (0.384) | (0.413) | (0.340) | (0.361) | (0.368) | (0.392) |
| *Time Served* | | -0.055 | | 0.075 | | 0.002 | | 0.0115 | | -0.138*** |
| | | (0.060) | | (0.079) | | (0.038) | | (0.043) | | (0.049) |
| *Out-group*Time Served* | | -0.105 | | -0.047 | | -0.136** | | -0.013 | | 0.011 |
| | | (0.084) | | (0.099) | | (0.059) | | (0.050) | | (0.054) |
| *Penalty* | | -0.033 | | 0.087** | | -0.011 | | 0.007 | | -0.027 |
| | | (0.032) | | (0.042) | | (0.014) | | (0.017) | | (0.021) |
| *High Security* | | -0.583 | | -3.062* | | 0.352 | | -0.408 | | 1.694* |
| | | (0.962) | | (1.584) | | (0.554) | | (0.713) | | (0.988) |
| *Married* | | 0.182 | | -0.721 | | 0.446* | | 0.231 | | -0.165 |
| | | (0.374) | | (0.616) | | (0.265) | | (0.295) | | (0.342) |
| *Number of Cell Mates* | | -0.338*** | | -0.056 | | -0.099 | | 0.088 | | 0.115 |
| | | (0.131) | | (0.147) | | (0.075) | | (0.090) | | (0.105) |
| *Age* | | 0.014 | | -0.041 | | -0.021 | | 0.035** | | -0.014 |
| | | (0.024) | | (0.025) | | (0.015) | | (0.017) | | (0.019) |
| *Education* | | -0.154 | | -0.308 | | -0.0183 | | -0.012 | | -0.372** |
| | | (0.187) | | (0.276) | | (0.134) | | (0.152) | | (0.180) |
| *Number of Children* | | 0.108 | | 0.258 | | 0.0513 | | -0.208** | | -0.035 |
| | | (0.116) | | (0.268) | | (0.099) | | (0.104) | | (0.116) |
| *Number of Siblings* | | 0.065* | | -0.092 | | -0.034 | | 0.043 | | -0.025 |
| | | (0.039) | | (0.073) | | (0.030) | | (0.032) | | (0.033) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Drug Related Offences* | | -0.592 | | 0.239 | | 0.610** | | -0.220 | | 0.657** |
| | | (0.384) | | (0.421) | | (0.240) | | (0.278) | | (0.328) |
| *Smuggling Goods* | | | | 1.101* | | -0.109 | | -0.758 | | 2.105** |
| | | | | (0.637) | | (0.466) | | (0.562) | | (0.879) |
| *Human Trafficking* | | -0.197 | | -0.685 | | 0.650* | | -0.536 | | 0.233 |
| | | (0.526) | | (0.815) | | (0.357) | | (0.425) | | (0.501) |
| *Homicide* | | 3.224 | | -3.776 | | 0.951 | | 0.236 | | 1.391 |
| | | (2.113) | | (2.670) | | (0.866) | | (1.003) | | (1.175) |
| *Sexual Crime* | | -0.439 | | 0.768 | | 0.0240 | | 0.657 | | 0.166 |
| | | (0.821) | | (1.374) | | (0.571) | | (0.689) | | (0.745) |
| *Economic Crime* | | 0.589 | | 0.378 | | 0.206 | | -1.283** | | 0.593 |
| | | (0.558) | | (0.788) | | (0.398) | | (0.547) | | (0.564) |
| *Firearms violations* | | 0.825 | | | | -0.187 | | -1.024* | | 0.590 |
| | | (0.643) | | | | (0.496) | | (0.578) | | (0.658) |
| *N* | 118 | 100 | 92 | 82 | 210 | 186 | 210 | 186 | 210 | 186 |
| *Priming + Out-group* Priming* | 1.213*** | 1.159** | 1.020** | 0.945* | 0.496* | 0.520 | 0.066 | -0.026 | 0.363 | 0.556 |
| *Out-group + Out-group* Priming* | 0.340 | 0.948** | -0.458 | -0.471 | -0.293 | 0.263 | 0.272 | 0.313 | 0.01 | -0.141 |
| *Time Served + Out-group*Time Served* | n/a | -0.159** | n/a | 0.029 | n/a | -0.134*** | n/a | -0.001 | n/a | -0.126*** |

**Notes:** Columns (1)-(6) present estimates from Probit regressions; columns (7)-(10) present estimates from ordered Probit regressions. Standard errors in parentheses. All specifications include subject random effects. Independent variables as defined in the text and in footnote 16. Last three rows present *p* values from tests on the restriction that the respective joint coefficient is equal to zero. *N* varies across specifications depending on the number of available observations (59 inmates as first movers, 46 inmates as second movers in the TG; data on control variables for some inmates missing in full specifications). Coefficient on *Smuggling Goods* missing in (2) since there were only two inmates in the role of the trustor and with that particular offence, leading this variable to perfectly predict success in the Probit model. *, **, *** indicates statistical significance at the 10%, 5%, 1% level, respectively.

**Table A3: Two-stage least squares estimation**

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | **Trust** | **Reciprocity** | **Cooperation** | **x-score** | **y-score** |
| *Out-group* | -0.041 | -0.657*** | -0.270** | 0.451 | -0.693 |
| | (0.164) | (0.195) | (0.129) | (0.583) | (0.595) |
| *Priming* | -0.003 | 0.017 | -0.152 | -0.555 | 0.934** |
| | (0.130) | (0.159) | (0.097) | (0.438) | (0.447) |
| *Out-group*Priming* | 0.369** | 0.249 | 0.303** | 0.434 | -0.226 |
| | (0.165) | (0.210) | (0.130) | (0.589) | (0.602) |
| *Time Served* | 0.004 | -0.075 | -0.011 | 0.181 | -0.435*** |
| | (0.029) | (0.062) | (0.025) | (0.115) | (0.117) |
| *Out-group*Time Served* | -0.042 | 0.070 | -0.027 | -0.160 | 0.227* |
| | (0.028) | (0.055) | (0.026) | (0.119) | (0.121) |
| *Penalty* | -0.006 | 0.0160* | -0.004 | 0.009 | -0.033 |
| | (0.008) | (0.008) | (0.005) | (0.022) | (0.022) |
| *High Security* | -0.188 | -0.760** | 0.130 | -0.250 | 1.352 |
| | (0.277) | (0.316) | (0.186) | (0.841) | (0.859) |
| *Married* | 0.068 | -0.345* | 0.137 | 0.420 | -0.369 |
| | (0.115) | (0.207) | (0.085) | (0.385) | (0.393) |
| *Number of Cell Mates* | -0.095*** | -0.034 | -0.030 | 0.125 | 0.132 |
| | (0.036) | (0.045) | (0.024) | (0.111) | (0.113) |
| *Age* | 0.002 | -0.004 | -0.006 | 0.050** | -0.011 |
| | (0.007) | (0.009) | (0.005) | (0.021) | (0.021) |
| *Education* | -0.064 | -0.127 | -0.003 | -0.025 | -0.340* |
| | (0.059) | (0.084) | (0.042) | (0.189) | (0.193) |
| *Number of Children* | 0.036 | 0.108 | 0.021 | -0.297** | -0.066 |
| | (0.034) | (0.081) | (0.029) | (0.131) | (0.134) |
| *Number of Siblings* | 0.017* | -0.022 | -0.009 | 0.045 | -0.022 |
| | (0.010) | (0.021) | (0.008) | (0.037) | (0.038) |
| *Drug Related Offences* | -0.167 | 0.126 | 0.212*** | -0.271 | 0.733** |
| | (0.105) | (0.135) | (0.076) | (0.346) | (0.353) |
| *Smuggling Goods* | 0.839** | 0.280 | -0.060 | -0.718 | 1.268* |
| | (0.387) | (0.205) | (0.149) | (0.673) | (0.687) |
| *Human Trafficking* | -0.112 | -0.145 | 0.230** | -0.734 | 0.480 |
| | (0.147) | (0.243) | (0.112) | (0.506) | (0.517) |
| *Homicide* | 0.704 | -0.769* | 0.336 | 0.434 | 2.025 |
| | (0.494) | (0.402) | (0.291) | (1.320) | (1.348) |
| *Sexual Crime* | -0.203 | 0.114 | -0.018 | 1.051 | 0.220 |
| | (0.254) | (0.421) | (0.194) | (0.880) | (0.899) |
| *Economic Crime* | 0.182 | 0.244 | 0.0834 | -1.584*** | 0.672 |
| | (0.167) | (0.264) | (0.127) | (0.577) | (0.589) |
| *Firearms violations* | 0.249 | -0.387 | -0.004 | -1.650** | 1.374* |
| | (0.190) | (0.387) | (0.155) | (0.700) | (0.715) |

| | | | | | |
|---|---|---|---|---|---|
| *N* | 102 | 84 | 186 | 186 | 186 |
| *Priming + Out-group* Priming* | 0.366*** | 0.266* | 0.151 | -0.121 | 0.708 |
| *Out-group + Out-group* Priming* | 0.328** | -0.408* | 0.033 | 0.885 | -0.919 |
| *Time Served + Out-group*Time Served* | -0.038** | -0.005 | -0.038*** | 0.021 | -0.208*** |
| *Sargan's Test of overidentifying restrictions* | *p*=0.360 | *p*=0.382 | *p*=0.138 | *p*=0.931 | *p*=0.904 |
| *Wu-Hausman test of endogeneity* | *p*=0.386 | *p*=0.105 | *p*=0.637 | *p*=0.157 | *p*=0.020 |

**Notes:** Instrumented variable: *Time Served*. Used instruments: frequency of leaving prison (days per month), working days per month. Standard errors in parantheses. Independent variables as defined in the text and in footnote 16. Last three rows present *p* values from tests on the restriction that the respective joint coefficient is equal to zero. N varies across specifications depending on the number of available observations (59 inmates as first movers, 46 inmates as second movers in the TG; data on control variables and instruments for some inmates missing). *, **, *** indicates statistical significance at the 10%, 5%, 1% level, respectively.

**Table A4: Distribution of crime categories in our sample and among U.S. prison population** (Source: Federal Bureau of Prisons)

| **Our sample** | | **U.S. prison population** | |
|---|---|---|---|
| Category | % of total | | % of total |
| Drug related | 35.0 | Drug offenses | 46.3 |
| Robbery | 29.3 | Burglary, Larceny, Property Offenses, Robbery | 8.5 |
| Homicide | 5.7 | Homicide, Aggravated Assault, and Kidnapping Offenses | 3.2 |
| Sexual crime | 2.4 | Sex offenses | 9.1 |
| Economic crime + smuggling goods | 11.4 | Banking and Insurance, Counterfeit, Embezzlement, Extortion, Fraud, Bribery | 6.7 |
| Firearms violations | 5.7 | Weapons, Explosives, Arson | 17.2 |
| Others (Smuggling goods, human trafficking) | 10.6 | Continuing Criminal Enterprise, Courts or corrections, Immigrations, National Security, Miscellaneous | 9.0 |

## 3. Decisions among students, in-group vs. out-group

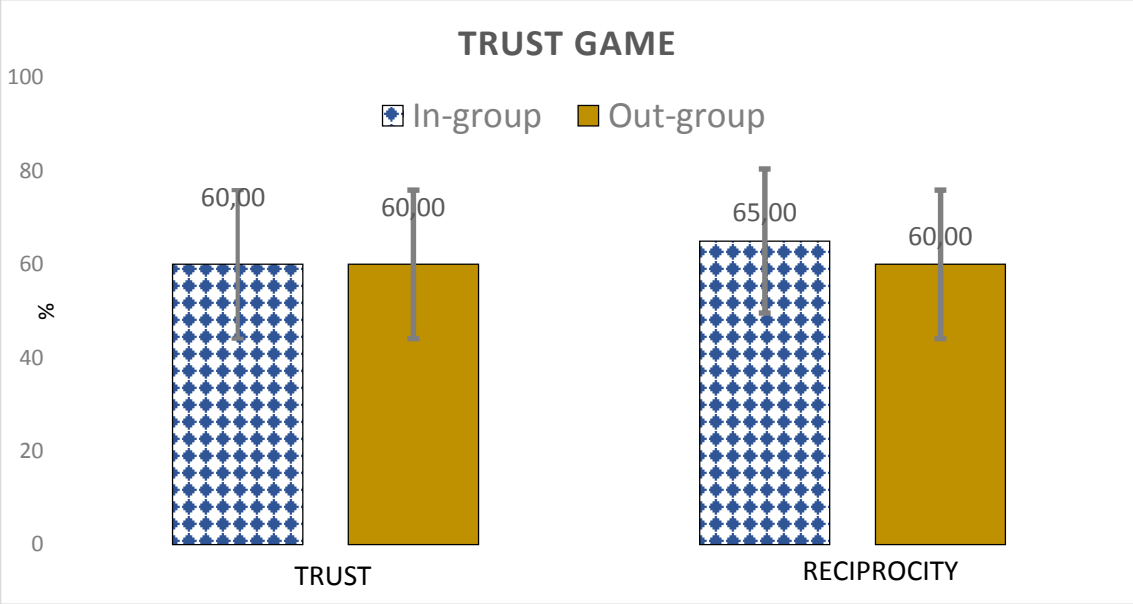


**Figure A4:** Decisions in the TG taken by students. Trust is measured as the percentage of first movers who choose to trust in the TG; reciprocity is measured as the percentage of second movers who choose to reciprocate in the TG. Bars include 95% confidence intervals.

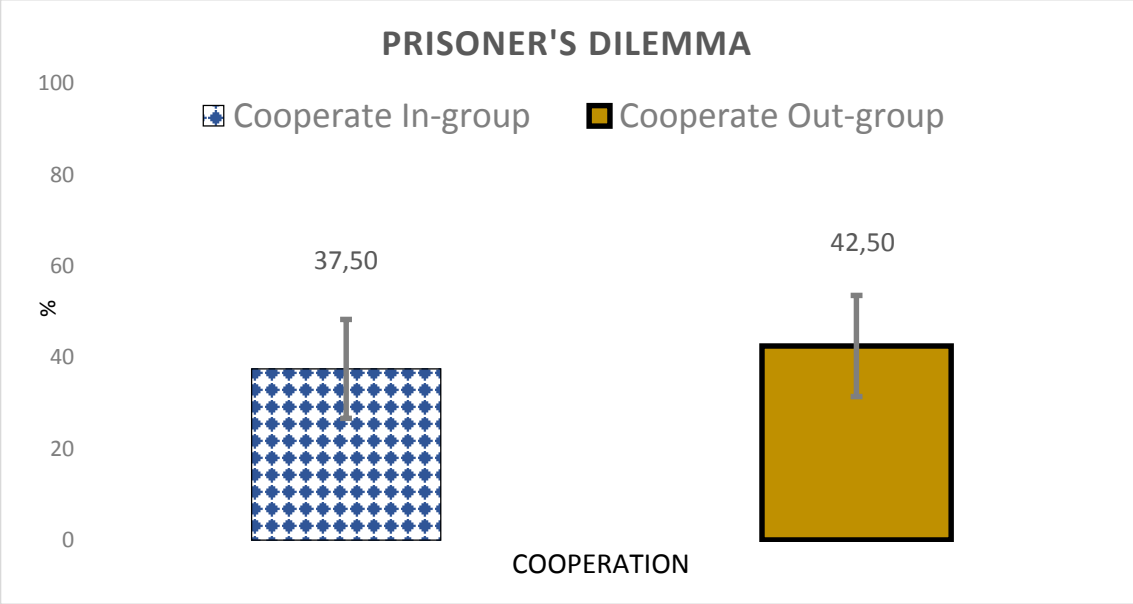


**Figure A5:** Decisions in the PD taken by students. Bars indicate cooperation rates and include 95% confidence intervals.

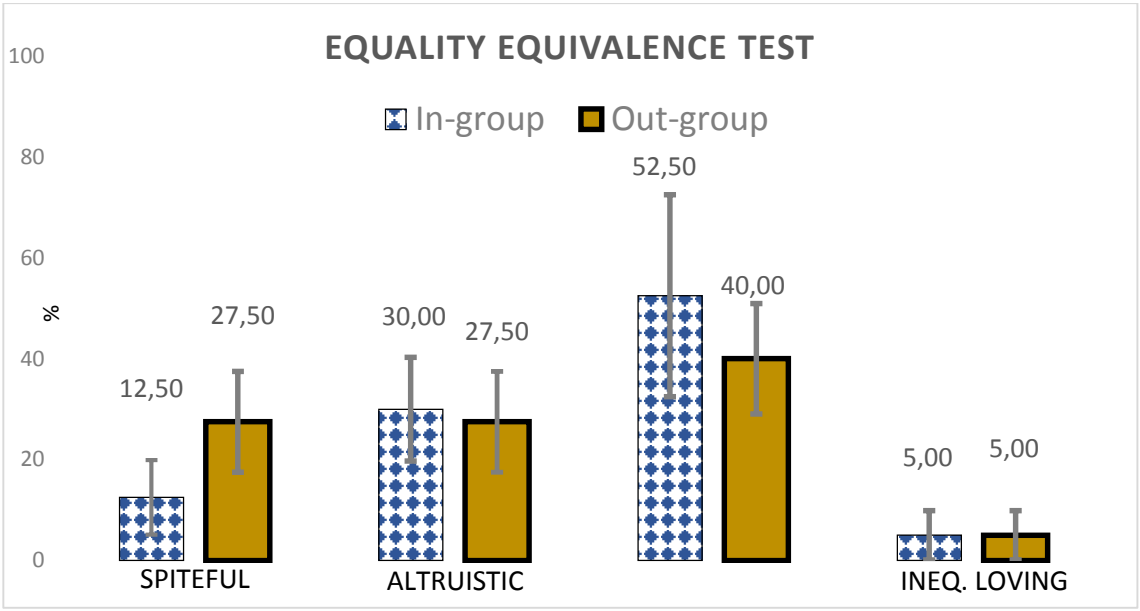


**Figure A6:** Classification of students' types (Spiteful, Altruistic, Inequality Averse, Inequality Loving) in the EET. Bars show the percentage of subjects classified into each type and include 95% confidence intervals.